W203 Lab 2

Student: Dili Wang | Section: 0904

1. **Meanwhile, at the Unfair Coin Factory:**

We are given the following:
- A bucket contains 99 fair coins
- The same bucket contains 1 trick coin which always comes up heads
- T is the event of selecting a trick coin. $P(T) = 0.01$

(a) Let H_k be the event that a coin comes up heads all k flips. To find the conditional probability that I have a trick coin given H_k, or $P(T|H_k)$, I can use Bayes Theorem:

$$P(T|H_k) = \frac{P(H_k|T) \cdot P(T)}{P(H_k)}$$

Next, I will find the terms for the right side of this equation:

i. $P(H_k|T) = 1$. Because the trick coin always comes up heads on every single flip, it will come up heads k times for k flips. Therefore the event H_k, which represents coming up heads k times for k flips, will always occur. Therefore, given that I have the trick coin, the probability of H_k is 100%, and thus $P(H_k|T) = 1$

ii. $P(T) = 0.01$. This is given in the problem.

iii. I can find $P(H_k)$ using the Law of Total Probability, which states that for mutually exclusive and exhaustive events $A_1, A_2, ...A_n$, an event B in the same event space can be expressed as:

$$P(B) = P(B|A_1) \cdot P(A_1) + P(B|A_2) \cdot P(A_2) + ... + P(B|A_n) \cdot P(A_n) = \sum_{i=1}^{n} P(B|A_i) \cdot P(A_i)$$

For the event space described by coins in the bucket, I can use T and $!T$ as mutually exclusive and exhaustive events. T is the event of picking the trick coin in a single draw, and $!T$ is then the event of picking a regular coin in a single draw. Since a single draw can only select a trick coin or a regular coin but not both, T and $!T$ are mutually exclusive. In addition, given that since there are 99 fair coins in the bucket, $P(!T) = \frac{99}{100} = 0.99$. $P(T) + P(!T) = 0.99 + 0.01 = 1$, so T and $!T$ fully partition the event space and are thus exhaustive. Using the Law of Total Probability, I can express $P(H_k)$ as:

$$\begin{aligned} P(H_k) &= P(H_k|T) \cdot P(T) + P(H_k|!T) \cdot P(!T) \\ &= 1 \cdot 0.01 + P(H_k|!T) \cdot 0.99 \end{aligned}$$

Lastly, to find $P(H_k|!T)$, I can use the fact that the probability of coming up heads for a single flip of a fair coin is $P(H_1|!T) = \frac{1}{2}$. Since k flips of a single coin are k independent events, I can multiply the probability of coming up heads for 1 flip k times:

$$\begin{aligned} P(H_k|!T) &= \underbrace{P(H_1|!T) \times P(H_1|!T) \times P(H_1|!T) \times ....}_{k\,times} \\ &= (P(H_1|!T))^k = \left[\frac{1}{2}\right]^k \end{aligned}$$

I then have the complete expression for $P(H_k)$,which is:

$$\begin{aligned} P(H_k) &= 1 \cdot 0.01 + P(H_k|!T) \cdot 0.99 \\ &= 0.01 + 0.99 \left[\frac{1}{2}\right]^k \end{aligned}$$

Collecting the terms from parts (i) through (iii), I have that the conditional probability of selecting a trick coin given the event H_k of flipping k heads in k flips is:

$$\begin{aligned} P(T|H_k) &= \frac{P(H_k|T) \cdot P(T)}{P(H_k)} \\ &= \frac{1 \cdot 0.01}{0.01 + 0.99\left[\frac{1}{2}\right]^k} \\ P(T|H_k) &= \boxed{\frac{0.01}{0.01 + 0.99\left[\frac{1}{2}\right]^k}} \end{aligned}$$

(b) To find how many heads in a row I would need to observe in order for the conditional probability of having a trick coin to be higher than 99%, I algebraically solve for k such that $P(T|H_k) > 0.99$:

$$\begin{aligned} 0.99 &< \frac{0.01}{0.01 + 0.99\left[\frac{1}{2}\right]^k} \\ 0.99 \cdot \left\{0.01 + 0.99\left[\frac{1}{2}\right]^k\right\} &< 0.01 \\ 0.0099 + 0.9801\left[\frac{1}{2}\right]^k &< 0.01 \\ \left[\frac{1}{2}\right]^k &< \frac{0.01 - 0.0099}{0.9801} \\ \left[\frac{1}{2}\right]^k &< \frac{0.0001}{0.9801} \\ ln\left(\left[\frac{1}{2}\right]^k\right) &< ln\left(\frac{0.0001}{0.9801}\right) \\ kln(\frac{1}{2}) &< ln\left(\frac{0.0001}{0.9801}\right) \\ k \cdot (-0.69315) &< -9.19024 \end{aligned}$$

To solve for k, I need to divide both sides by $ln(2) \approx -0.69315$. Dividing by a negative quantity will switch the inequality sign, so the result is:

$$\begin{aligned} k \cdot (-0.69315) &< -9.19024 \\ k &> \frac{-9.19024}{-0.69315} \\ \rightarrow k &> \frac{ln\left(\frac{0.0001}{0.9801}\right)}{ln\left(\frac{1}{2}\right)} \\ \rightarrow k &> 13.26 \end{aligned}$$

Since k must be an integer quantity, given that it is the number of flips, the result is that $\boxed{k \geq 14}$ flips in order for the probability $P(T|H_k) > 0.99$

2. **Wise Investments:**

We are given the following:
- X be a random variable of the total number of companies that reach unicorn status. X can take on the values 0, 1 or 2
- The probability of a company reaching unicorn status is $p = \frac{3}{4}$, independent of the other company

(a) To find the probability mass function of X, I need to find the probability for each individual value of X. Here, I will use U_i to denote the event that a company i achieves unicorn status, where $i$takes on the value of 1 or 2, since I have 2 companies. I am given that $P(U_i) = \frac{3}{4}$, for any i, which would mean that the probability of any company i does not achieve unicorn status is $P(!U_i) = 1 - \frac{3}{4} = \frac{1}{4}$. I can then find the probability mass function by taking the following steps:

i. $X = 0$ means that neither company achieves unicorn status. Because the event of either company achieving unicorn status is independent of the other company, the event of either company not achieving unicorn status is also independent of the other. I can then find the probability of neither company achieving unicorn status by multiplying their individual probabilities $(P(!U_1), P(!U_2))$ of non-unicorn status together:

$$P(X = 0) = P(!U_1) \cdot P(!U_2) = \frac{1}{4} \times \frac{1}{4} = \frac{1}{16}$$

ii. $X = 1$ means that one of the 2 companies have achieved unicorn status. However, this can be accomplished in one of 2 ways, either company 1 achieves it and company 2 does not $(P(U_1) \cdot P(!U_2))$ or company 2 achieves it and company 1 does not $(P(!U_1) \cdot P(U_2))$. Again, since either company achieving or not achieving unicorn status are independent events, I can multiple the individual events probabilities together. The total probability for $X = 1$, then the sum of the probability of the 2 ways in which one of the companies will achieve unicorn status:

$$P(X = 1) = P(U_1) \cdot P(!U_2) + P(!U_1) \cdot P(U_2) = \frac{3}{4} \times \frac{1}{4} + \frac{1}{4} \times \frac{3}{4} = \frac{6}{16} = \frac{3}{8}$$

iii. $X = 2$ means that both companies have achieved unicorn status. Again, using the independence of the 2 events of either company achieving unicorn status, I can calculated $P(X = 2)$ as follows:

$$P(X = 2) = P(U_1)P(U_2) = \frac{3}{4} \times \frac{3}{4} = \frac{9}{16}$$

I can then express the probability mass function, $p(x) = P(X = x)$, of the discrete random variable X as:

$$\boxed{p(x)} = \begin{cases} \frac{1}{16} & if\ x = 0 \\ \frac{3}{8} & if\ x = 1 \\ \frac{9}{16} & if\ x = 2 \end{cases}$$

(b) The cumulative probability function, $F(x)$ of a discrete random variable X with pmf $p(x)$ is defined as:

$$F(x) = P(X \leq x) = \sum_{y:y \leq x} p(y)$$

Therefore, in order to find the value for $F(x)$ for each value of x, I need to sum the values of $p(x)$ up to and including that value of x:

i. $F(0) = P(X \leq 0) = P(X = 0) = p(x = 0) = \frac{1}{16}$
This value of $F(x)$ exists in the region of $0 \leq x < 1$

ii. $F(1) = P(X \leq 1) = P(X = 0\, or\, 1) = (p(x = 0) + p(x = 1) = \frac{1}{16} + \frac{3}{8} = \frac{7}{16}$
This value of $F(x)$ exists in the region of $1 \leq x < 2$

iii. $F(2) = P(X \leq 2) = P(X = 0\, or\, 1\, or\, 2) = p(x=0) + p(x=1) + p(x=2) = \frac{1}{16} + \frac{3}{8} + \frac{9}{16} = 1$

This value of $F(x)$ exists in the region of $x \geq 2$

Lastly, I can conclude that $F(x < 0) = 0$, since there is no probability mass function defined for $x < 0$. Collectively, I can write the cumulative probability function of X as:

$$\boxed{F(x)} = \begin{cases} 0 & x < 0 \\ \frac{1}{16} & 0 \leq x < 1 \\ \frac{7}{16} & 1 \leq x < 2 \\ 1 & x \geq 2 \end{cases}$$

(c) For the discrete random variable X, the expected value, $E(X)$, is defined as:

$$E(X) = \sum_{all\, possible\, x} x \cdot p(x)$$

In this problem, the possible values of $x$are $x_0 = 0$, $x_1 = 1$, and $x_2 = 2$. I can then write $E(X)$ as:

$$\begin{aligned} E(X) &= \sum_{i=0}^{2} x_i \cdot p(x_i) \\ &= x_0 \cdot p(x_0) + x_1 \cdot p(x_1) + x_2 \cdot p(x_2) \\ &= 0 \cdot p(x=0) + 1 \cdot p(x=1) + 2 \cdot p(x=2) \\ &= 0 + 1 \times \frac{3}{8} + 2 \times \frac{9}{16} \\ E(X) &= \boxed{1.5} \end{aligned}$$

(d) The variance of a discrete random variable X, $var(X)$, is defined using the expected value $\mu = E(X)$ as:

$$var(X) = \sum_{all\, possible\, x} (x - \mu)^2 \cdot p(x)$$

Using the same set of possible $x$values as (c), I can write $var(X)$ as:

$$\begin{aligned} var(X) &= \sum_{i=0}^{2} (x_i - E(X))^2 \cdot p(x_i) \\ &= (x_0 - 1.5)^2 \cdot p(x_0) + (x_1 - 1.5)^2 \cdot p(x_1) + (x_2 - 1.5)^2 \cdot p(x_2) \\ &= (0 - 1.5)^2 \cdot p(0) + (1 - 1.5)^2 \cdot p(1) + (2 - 1.5)^2 \cdot p(2) \\ &= 2.25 \times \frac{1}{16} + 0.25 \times \frac{3}{8} + 0.25 \times \frac{9}{16} \\ var(X) &= \boxed{\frac{3}{8}} \end{aligned}$$

3. **Relating Min and Max**

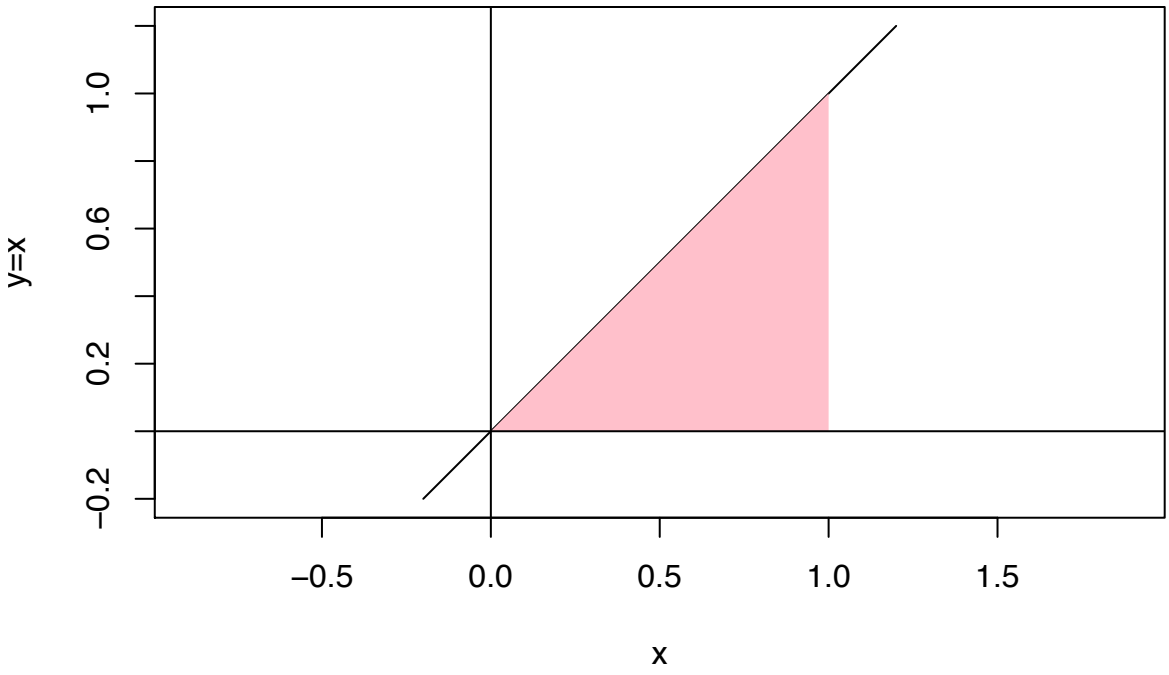
We are given the following:
For continuous random variables X and Y, the joint probability density function is defined as:

$$f(x,y) = \begin{cases} 2 & 0 < y < x < 1 \\ 0 & otherwise \end{cases}$$

(a) For a graph of the region for which X and Y have positive probability, can first plot a line $y = x$. The shaded triangular region under the line represents the area where y<x, and the shaded region is limited to $0 < x < 1$. This is then the area where the probability is positive.

```
x = seq(-0.2,1.2,length=100)
y = x


plot(x, y, type="l", ylab=expression("y=x"), asp = 1)
polygon(c(0:1, 1:1), c(0:1, 0:0), col= "pink", border=NA)
abline(h=0, v=0)
```



(b) The marginal probability density function of X is $f_X(x)$ and this is defined as:

$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)dy \qquad for -\infty < x < \infty$$

Here, the limits of integration with respect to y is from 0 to x, since $f(x,y)$ is positive only in the region of $0 < y < x$ and $0 < x < 1$. I then have the following calculation for $f_X(x)$:

$$\begin{aligned} f_X(x) &= \int_0^x f(x,y)dy \\ &= \int_0^x 2dy \\ &= 2y|_0^x = 2x - 2 \cdot 0 = 2x \end{aligned}$$

$$\rightarrow \boxed{f_X(x)} = \begin{cases} 2x & 0 < x < 1 \\ 0 & elsewhere \end{cases}$$

(c) The unconditional expectation of X is defined as follows:

$$\begin{aligned} E(X) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x \cdot f(x,y)dxdy \\ &= \int_{-\infty}^{\infty} x \left[\int_{-\infty}^{\infty} f(x,y)dy \right] dx \\ &= \int_{-\infty}^{\infty} x \cdot f_X(x)dx \end{aligned}$$

Using the fact that $f_X(x)$ is non-zero for only the region $0 < x < 1$, the integral with respect to x is limited from 0 to 1:

$$\begin{aligned} E(X) &= \int_0^1 x \cdot f_X(x)dx \\ &= \int_0^1 x \cdot 2xdx \\ &= \int_0^1 2x^2 dx \\ &= 2 \cdot \frac{x^3}{3}\Big|_0^1 = 2 \times \frac{1^3}{3} - 2 \times \frac{0^3}{3} \\ E(X) &= \boxed{\frac{2}{3}} \end{aligned}$$

(d) The conditional probability density function of Y, conditioned on X, can be derived using the definition, $f_{Y|X}(y|x) = \dfrac{f(x,y)}{f_X(x)}$. I am already given the numerator in this equation, and the denominator was derived in (b). Keeping in mind that $f(x,y)$ is positive only in the region $0 < y < x$ and $0 < x < 1$, conditional probability density function is derived as follows:

$$f_{Y|X}(y|x) = \frac{f(x,y)}{f_X(x)}$$
$$= \frac{2}{2x}$$
$$= \frac{1}{x}$$

$$\boxed{f_{Y|X}(y|x)} = \begin{cases} \frac{1}{x} & 0 < y < x < 1 \\ 0 & elsewhere \end{cases}$$

(e) The conditional expectation of Y, conditional on X, is the expectation of Y using the conditional probability distribution function $f_{Y|X}(y|x)$. In other words, we derive $E(Y|X) = E(Y|X = x)$, using the following integral with respect to y while treating x as if it were a constant with respect to y:

$$E(Y|X) = \int_{-\infty}^{\infty} y \cdot f_{Y|X}(y|x)dy$$

given that $f_{Y|X}(y|x)$ is only positive and non-zero for the region $0 < y < x$, we will need to limit the integration with respect to $y$from 0 to x:

$$E(Y|X) = \int_0^x y \cdot f_{Y|X}(y|x)dy$$
$$= \int_0^x y \cdot \frac{1}{x}dy$$
$$= \frac{y^2}{2x}\Big|_0^x = \frac{x^2}{2x} - 0$$
$$E(Y|X) = \boxed{\frac{x}{2}}$$

(f) In order to derive $E(XY)$, I can first use the Law of Iterated Expectations which states that $E(XY) = E(E(XY|X))$. Given the fact that the X can be treated as a constant inside of an expectation that is conditional on X, I can use $E(XY|X) = XE(Y|X)$. I then have the following:

$$E(XY) = E(E(XY|X))$$
$$= E(XE(Y|X))$$

For any function, $g(x)$, the expectation $E(g(x))$ can be derived as follows:

$$E(g(x)) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x) \cdot f(x,y)dxdy$$
$$= \int_{-\infty}^{\infty} g(x)\left[\int_{-\infty}^{\infty} f(x,y)dy\right]dx$$
$$= \int_{-\infty}^{\infty} g(x) \cdot f_X(x)dx$$

I can then derive $E(XY) = E(XE(Y|X))$ using the fact that this is the expectation of some function of x. Substituting $g(x) = xE(Y|X)$ into the integral for $E(g(x))$ defined above, I can solve for $E(XY)$ using this integral:

$$E(XY) = E(XE(Y|X)) = \int_{-\infty}^{\infty} x \cdot E(Y|X) \cdot f_X(x)dx$$

Using the result from (e), I can substitute $E(Y|X) = \frac{x}{2}$ and $f_X(x) = 2x$ into the above equation, and set the limits of integration from 0 to 1, since $f_X(x)$ is non-zero and positive only for $0 < x < 1$:

$$\begin{aligned} E(XY) = E(XE(Y|X)) &= \int_0^1 x \cdot E(Y|X) \cdot f_X(x)dx \\ &= \int_0^1 x \cdot \frac{x}{2} \cdot 2xdx \\ &= \int_0^1 x^3 dx \\ &= \frac{x^4}{4}\Big|_0^1 = \frac{1}{4} - 0 \\ E(XY) &= \boxed{\frac{1}{4}} \end{aligned}$$

(g) The covariance of X and Y is defined as $cov(X,Y) = E(XY) - E(X)E(Y)$. I have all of the terms needed for the covariance from parts (c) and (f), with the exception of $E(Y)$.
To derive $E(Y)$, I can first use the Law of Iterated Expectations, which states that $E(Y) = E(E(Y|X))$. From (e), I have that $E(Y|X) = \frac{x}{2}$, which is just another function of x. I can then use the definition of the expectation of a function derived in (f):

$$E(g(x)) = \int_{-\infty}^{\infty} g(x) \cdot f_X(x)dx$$

except this time, I make a substitution $g(x) = \frac{x}{2}$ and $f_X(x) = 2x$. Again, limiting the integral for 0 to 1 given that $f_X(x)$ is only non-zero and positive for $0 < x < 1$, I can derive $E(Y)$ as follows:

$$\begin{aligned} E(Y) = E(E(Y|X)) &= E(\frac{x}{2}) \\ &= \int_0^1 \frac{x}{2} \cdot f_X(x)dx \\ &= \int_0^1 \frac{x}{2} \cdot 2xdx \\ &= \int_0^1 x^2 dx \\ &= \frac{x^3}{3}\Big|_0^1 = \frac{1}{3} \end{aligned}$$

Using the result from above and the results from (c)$E(X) = \frac{2}{3}$ and (f)$E(XY) = \frac{1}{4}$, I can calculate the covariance:

$$\begin{aligned} cov(X,Y) &= E(XY) - E(X)E(Y) \\ &= \frac{1}{4} - \frac{2}{3} \times \frac{1}{3} \\ &= \frac{1}{4} - \frac{2}{9} = \frac{9}{36} - \frac{8}{36} \\ cov(X,Y) &= \boxed{\frac{1}{36}} \end{aligned}$$

4. **Circles, Random Samples, and the Central Limit Theorem**

We are given the following:
- $X_1, X_2, ..., X_n$ and $Y_1, Y_2, ..., Y_n$ are independent random samples from a uniform distribution on [-1,1].
- D_i is a random variable that indicates if (X_i, Y_i) falls within the unit circle centered at the origin. D_i is defined as:

$$D_i = \begin{cases} 1 & X_i^2 + Y_i^2 < 1 \\ 0 & otherwise \end{cases}$$

- all D_i are independent and identically distributed

(a) To compute $E(D_i)$ I can use it's definition:

$$E(D_i) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} D_i f(x,y) dx dy$$

To find the joint probability distribution function, we can use the given statement that X_i and Y_i are independent random samples with uniform distribution.
The marginal probability distribution of X_i, which is a uniform distribution on [-1, 1], is then defined as:

$$f_X(x) = \begin{cases} \dfrac{1}{1-(-1)} = \dfrac{1}{2} & -1 \le x \le 1 \\ 0 & elsewhere \end{cases} = \begin{cases} \dfrac{1}{2} & -1 \le x \le 1 \\ 0 & elsewhere \end{cases}$$

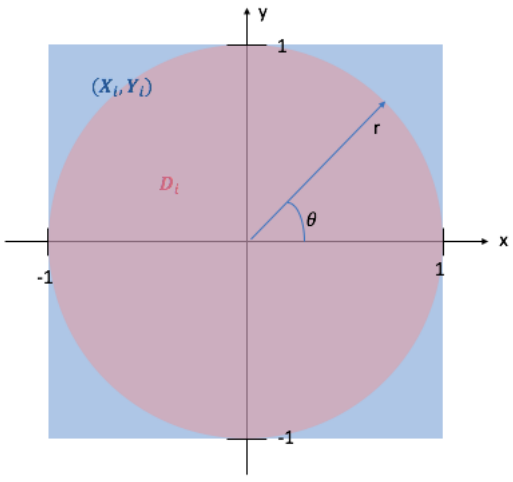
Similarly, the marginal probability distribution of Y_i, which is also uniformly distributed on [-1,1], is defined as:

$$f_Y(y) = \begin{cases} \dfrac{1}{2} & -1 \le y \le 1 \\ 0 & elsewhere \end{cases}$$

Given the independence of X_i and Y_i, we can use the definition on page 207 of Devore's *Probability and Statistics for Engineering and the Sciences*, which states that the joint pdf of 2 random variables is the product of their marginal pdfs. Therefore, I can compute $f(x,y)$ as:

$$\begin{aligned} f(x,y) &= f_X(x) \cdot f_Y(y) \\ &= \begin{cases} \dfrac{1}{2} \cdot \dfrac{1}{2} & -1 \le x \le 1 \, and \, -1 \le y \le 1 \\ 0 & elsewhere \end{cases} \\ f(x,y) &= \begin{cases} \dfrac{1}{4} & -1 \le x \le 1 \, and \, -1 \le y \le 1 \\ 0 & elsewhere \end{cases} \end{aligned}$$

With the distribution of (X_i, Y_i) and D_i, I constructed the following figure which I will refer to for the rest of my solution:



As shown in this figure, the joint probability distribution of (X_i, Y_i) is non-zero and positive in the region

of $-1 \leq x \leq 1$ and $-1 \leq y \leq 1$, and thus (X_i, Y_i) is distributed within a square of Area = 4 (shaded in blue). D_i is distributed within the pink circle of radius =1 that is enclosed within the square.

With $f(x, y)$ derived above, I can now compute the integral for $E(D_i)$. First, I will need limit the integral to -1 to 1 with respect to both $x$and y, given the limits the distribution of (X_i, Y_i):

$$\begin{aligned} E(D_i) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} D_i f(x, y) dxdy \\ &= \int_{-1}^{1} \int_{-1}^{1} D_i \cdot \frac{1}{4} dxdy \end{aligned}$$

Next, I will use the fact that $D_i = 1$ only the region of $X_i^2 + Y_i^2 < 1$, which further limits the integral to points (x, y) confined to the unit circle:

$$E(D_i) = \iint_{(x^2+y^2)<1} 1 \cdot \frac{1}{4} dxdy$$

For the integral of a function over the area of a circle of radius R, we can use the following coordinate transformation to polar coordinates (r, θ), where $r$is the distance from the origin $(0, 0)$ and θ is the angular coordinate measured counterclockwise from the positive x-axis, as shown in the above figure:

$$\iint_{(x^2+y^2)<R^2} h(x, y) dxdy = \int_{\theta=0}^{\theta=2\pi} \int_{0}^{R} h(r, \theta) \cdot r dr d\theta$$

Here, I am integrating over the area of a unit circle of $R = 1$ centered at the origin, and the integrand function is a constant: $h(x, y) = h(r, \theta) = \frac{1}{4}$. Transforming coordinates of integration from (x, y) to (r, θ), I can then re-write and calculate $E(D_i)$ as:

$$\begin{aligned} E(D_i) &= \iint_{(x^2+y^2)<1} 1 \cdot \frac{1}{4} dxdy \\ &= \int_{\theta=0}^{\theta=2\pi} \int_{0}^{1} \frac{1}{4} \cdot r dr d\theta \\ &= \int_{0}^{2\pi} \left[\frac{r^2}{8} \Big|_0^1 \right] d\theta = \int_{0}^{2\pi} \left[\frac{1}{8} - 0 \right] d\theta \\ &= \int_{0}^{2\pi} \frac{1}{8} d\theta = \frac{\theta}{8} \Big|_0^{2\pi} = \frac{2\pi}{8} - 0 \\ E(D_i) &= \boxed{\frac{\pi}{4}} \end{aligned}$$

To check my result above, I can also use an alternative approach to calculate $E(D_i)$, which considers it as a weighted average defined by:

$$\frac{sum\, of\, (all\, points\, (X_i, Y_i)\, where\, D_i = 1)}{sum\, of\, all\, points\, (X_i, Y_i)}$$

The numerator, $sum\, of\, (all\, points\, (X_i, Y_i)\, where\, D_i = 1)$ is all of the points (X_i, Y_i) in the unit circle, which is then just the Area of the Unit Circle. The denominator, $sum\, of\, all\, points\, (X_i, Y_i)$, is then all of the points (X_i, Y_i), whether or not they are in the unit circle, which is then the area of the blue square. $E(D_i)$ is then equal to $\frac{Area\, of\, Unit\, Circle}{Area\, of\, Square} = \frac{\pi \cdot 1^2}{4} = \frac{\pi}{4}$. This alternative calculation checks out with the result of my integration for $E(D_i)$ shown above.

(b) For the standard deviation of D_i, denoted as σ_{D_i}, I will be using the definition of variance: $var(D_i) = (\sigma_{D_i})^2$. First I will compute the variance of D_i, defined as $var(D_i) = E(D_i^2) - E(D_i)^2$, then take the

square root. I can use the result computed for $E(D_i)$ from (a), and then use the same method of integration over polar coordinates to compute $E(D_i)^2$:

$$\begin{aligned} var(D_i) &= E(D_i^2) - E(D_i)^2 \\ &= \iint_{(x^2+y^2)<1} D_i^2 \cdot \frac{1}{4} dxdy - E(D_i)^2 \\ &= \int_0^{2\pi} \int_0^1 1^2 \cdot \frac{1}{4} \cdot r dr d\theta - E(D_i)^2 \\ &= \int_0^{2\pi} \left[\frac{r^2}{8} \Big|_0^1 \right] d\theta - \left(\frac{\pi}{4}\right)^2 \\ &= \int_0^{2\pi} \frac{1}{8} d\theta - \left(\frac{\pi}{4}\right)^2 \\ &= \frac{\theta}{8} \Big|_0^{2\pi} - \left(\frac{\pi}{4}\right)^2 = \frac{\pi}{4} - \left(\frac{\pi}{4}\right)^2 \\ \sigma_{D_i} &= \sqrt{var(D_i)} = \boxed{\sqrt{\frac{\pi}{4} - \left(\frac{\pi}{4}\right)^2}} \approx \boxed{0.410546} \end{aligned}$$

(c) The standard error of $\bar{D}$ is defined as $\sigma_{\bar{D}} = \frac{\sigma_D}{\sqrt{n}}$. Using our result in (b):

$$\sigma_{\bar{D}} = \frac{\sigma_D}{\sqrt{n}} = \frac{\sqrt{\frac{\pi}{4} - \left(\frac{\pi}{4}\right)^2}}{\sqrt{n}} = \boxed{\sqrt{\frac{\frac{\pi}{4} - \left(\frac{\pi}{4}\right)^2}{n}}}$$

(d) The Central Limit Theorem states that for a sufficiently large random sample of size n, $D_1, D_2, ..., D_n$, selected from a distribution with mean μ_D and variance σ_D^2, $\bar{D}$ has approximately a normal distribution with $\mu_{\bar{D}} = \mu_D$ and $\sigma_{\bar{D}}^2 = \frac{\sigma_D^2}{n}$. Additionally, the larger the value of n, the better the approximation. (This definition is taken directly from page 232 of Devore's *Probability and Statistics for Engineering and the Sciences* and applied here to the problem's random variable D_i).

Since the sample size of $n = 100$ is considered sufficiently large, the Central Limit Theorem allows us to approximate the distribution of $\bar{D}$ as a normal distribution that has mean $\mu_{\bar{D}} = \mu_D$ and standard deviation equal to the standard error $\sigma_{\bar{D}}$.
The mean of the sampling distribution is just the expectation of D_i from (a):

$$\mu_{\bar{D}} = \mu_D = E(D) = \frac{\pi}{4}$$

The standard deviation of the sampling distribution is:

$$\sigma_{\bar{D}} = \sigma_D = \sqrt{\frac{\frac{\pi}{4} - \left(\frac{\pi}{4}\right)^2}{100}} \approx 0.0410546$$

With this information, I can then use the pnorm() function in R to calculate the probability that $\bar{D} > \frac{3}{4}$. The function pnorm() has the first 3 arguments - the critical value q, the mean, and the standard deviation sd. Using our normal approximation of the sampling distribution for $\bar{D}$ with mean = $\mu_{\bar{D}}$ and standard deviation = $\sigma_{\bar{D}}$, I can then use the following to compute $P(\bar{D} > \frac{3}{4})$:

$$\begin{aligned} P(\bar{D} \leq \frac{3}{4}) &= pnorm(\frac{3}{4}, mean = \mu_{\bar{D}}, sd = \sigma_{\bar{D}}) \\ P(\bar{D} > \frac{3}{4}) &= 1 - P(\bar{D} \leq \frac{3}{4}) = 1 - pnorm(\frac{3}{4}, mean = \mu_{\bar{D}}, sd = \sigma_{\bar{D}}) \end{aligned}$$

In my R code below, I use the following variable names:

- σ_D is represented by sigmaD, which is the standard deviation calculated from (b)
- $\sigma_{\bar{D}}$, the standard error, is represented by sigma_barD
- $\mu_{\bar{D}}$, the sample distribution mean, is represented by mu_barD

```
#storing population standard deviation
(sigmaD = ((pi/4)-(pi/4)**2)**0.5)
```

```
## [1] 0.4105458
```

```
#using standard deviation to caculate standard error
(sigma_barD = sigmaD/(100**0.5))
```

```
## [1] 0.04105458
```

```
#storing the sample distribution mean
(mu_barD = pi/4)
```

```
## [1] 0.7853982
```

```
#calculating the probability of sample mean greater than 3/4
(1-pnorm(3/4, mean = mu_barD, sd=sigma_barD))
```

```
## [1] 0.8057173
```

From this output, we have that the $P(\bar{D} > \frac{3}{4}) \approx \boxed{0.8057}$

(e) Simulating a single draw can be done using runif() which creates random samples for a uniformly distributed random variable within the minimum and maximum bounds used to define the distribution. The first 3 arguments of the runif() function are the number of samples n, the min bound of the distribution, and the max bound of the distribution.

As discussed in (a), X_i and Y_i have uniform distribution in the region [-1, 1], so the arguments for runif() can be set as follows:

- n=100
- min = -1
- max = 1

Given the independence of X_i and Y_i, runif() is ran separately to create the sample results stored in vectors x and y:

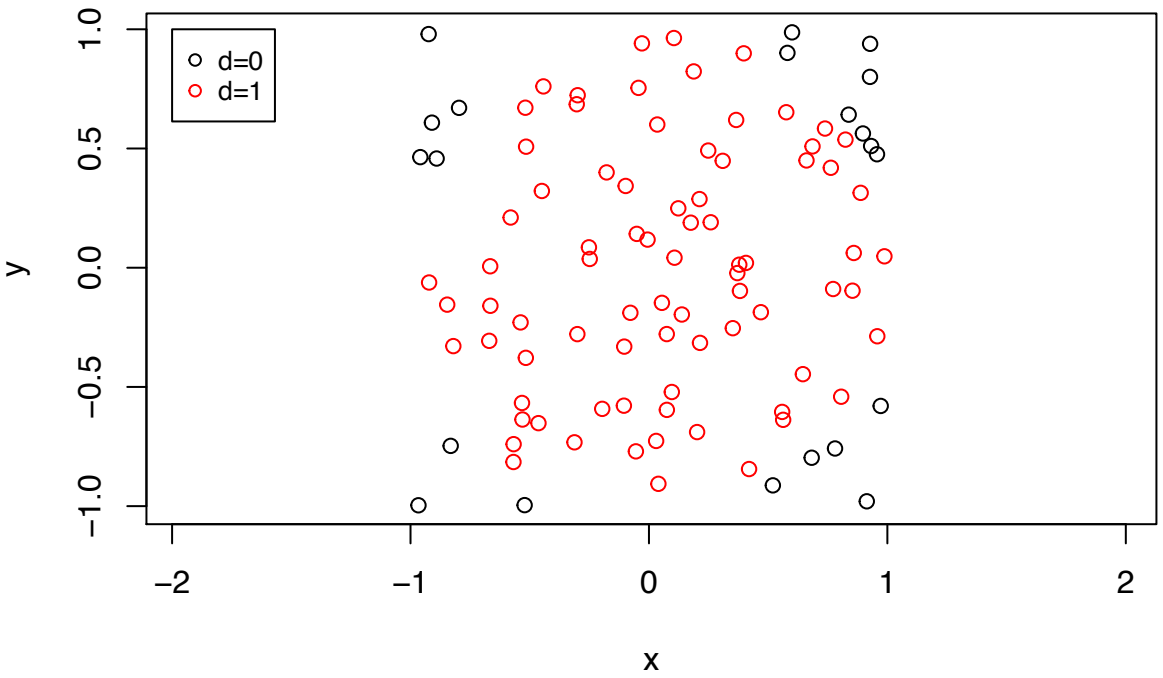
```
x = runif(100, min = -1, max = 1)
y = runif(100, min = -1, max = 1)
```

Next, I create a function called D_i for calculating D_i, which takes 2 arguments that represent X_i and Y_i. Then I store the results for D_i operating on the vectors x and y in the results vector d

```
D_i = function(x,y) {
  ifelse((x**2+y**2)<1, 1, 0)
}
d = D_i(x,y)
```

Finally, I plot my samples (x,y) in the x-y plane where the data points for $D_i = 1$ (which corresponds to d=1) are color-coded red, and data points for $D_i = 1$ are black:

```
plot(x,y, col=d+1, asp = 1, main = 'Distribution of X, Y and D')
legend(-2, 1, legend=c("d=0", "d=1"),
       col=c("black", "red"), cex=0.8, pch = 1:1)
```



From this graph, we can see that for n=100 random samples, the complete distribution of (x,y) rougly resembles a square that spans from -1 to 1 for x and from -1 to 1 for y. The red points that correspond to d = 1 roughly form the shape of the unit circle. As expected, the points of the sample lie within the square bounds and the points for d=1 lie within the circular bounds that correspond to the figure I created for part (a).

(f) The sample average, $\bar{D}$ can be computed using the mean() function as follows:

```
#sample average
(sample_mean = mean(d))
```

```
## [1] 0.79
```

```
#expected sample average from (a)
(pi/4)
```

```
## [1] 0.7853982
```

The two outputs above show that the average of the sample and the expected mean calculated from part (a) $\frac{\pi}{4} \approx 0.7853982$, are quite close, but not exactly equal to each other. This is expected, since the sample distribution for $\bar{D}$ has non-zero variance. We would expect the measurement of $\bar{D}$ for a sample to deviate from the expected mean and also to be different from sample to sample.

(g) To generate 10,000 samples, we can use the replicate() function in R. First, we will need to define a function, D_bar, which will accomplish the following:

- Create X and Y samples of size n.
- Calculated the corresponding value of D_i for every (X_i, Y_i) using the previously defined function D_i.
- Calculate and return the sample mean $\bar{D}$ as the output.

```
D_bar = function(n){
  x_j = runif(n, -1, 1)
  y_j = runif(n, -1, 1)
```

```
  d_j = D_i(x_j,y_j)
  return(mean(d_j))
}
```

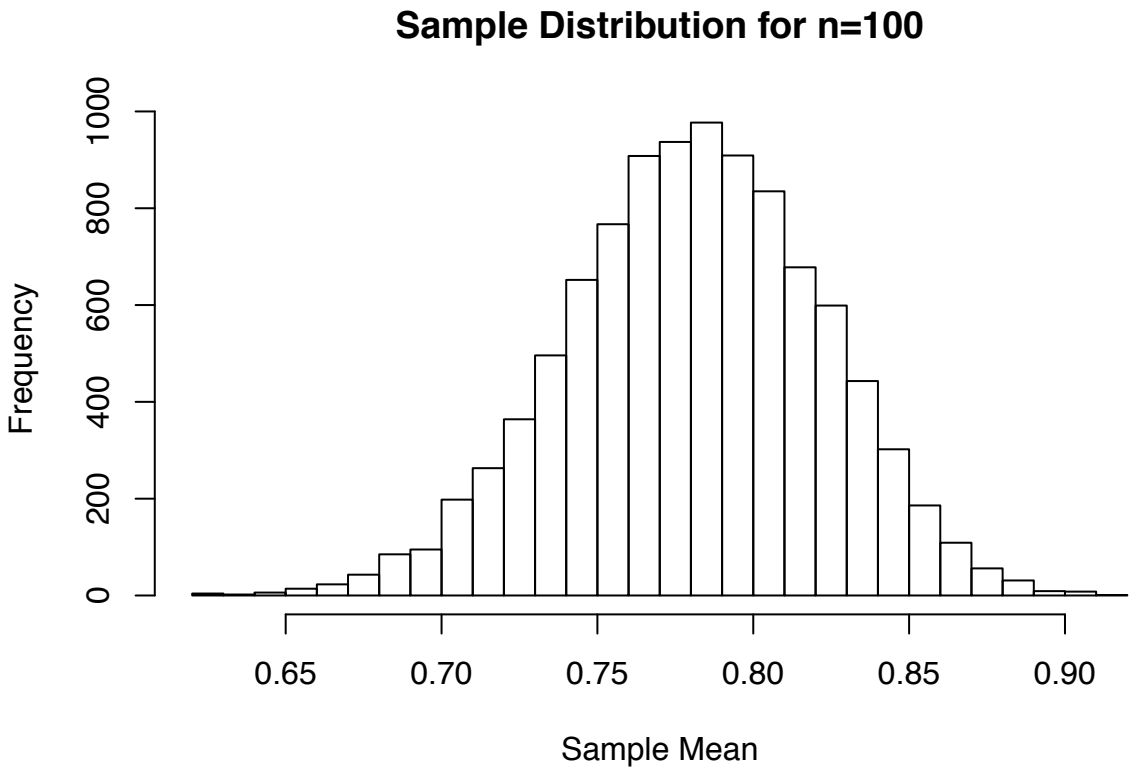
Next, I will set the sample size and replicate the simulation function D_bar 10,000 times:

```
#set sample size
n = 100

#store sample means for 10,000 draws in vector
sample_means_D = replicate(10000,  D_bar(n))
```

Lastly, I plot a histogram for the results of the simulation, which are 10,000 values of $\bar{D}$ stored in the vector sample_means_D:

```
# plot sample means as histogram
hist(sample_means_D, breaks=40, xlab = 'Sample Mean',
     ylab = 'Frequency', main = 'Sample Distribution for n=100')
```



From the histogram above, we see that the sample distribution for $\bar{D}$ has the rougly the shape of a normal distribution.

(h) The standard deviation of sample means can be calculated using the sd() function as follows:

```
sd(sample_means_D)
```

```
## [1] 0.04109532
```

The output here is also very close in comparison to the result from (c) 0.0410546. This result aligns with our expectation from the CLT, which states that for a sufficiently large sample, sample distribution can be approximated with a normal distribution that has standard deviation equal to the standard error found in (c)

(j) To calculate the fraction of sample averages that are larger than 3/4, I use length() function which

counts the number of elements. I can then slice sample_means_D to pick out a subset of numbers that are greater than 3/4, then divide the length() of this subset by the length of the entire vector() sample_means_D:

```
length(sample_means_D[sample_means_D>0.75])/length(sample_means_D)
```

```
## [1] 0.7755
```

This output from the simulation of 10,000 draws is very close to answer from (d) $P(\bar{D} > \frac{3}{4}) \approx 0.806$, but the two values are also not exactly equal.

From both the CLT and the histogram shown in (g), we know that the normal distribution is only an approximation for the sample distribution of $\bar{D}$, but there will be some error. Therefore we should expect some deviation of the actual observed fraction of samples where $\bar{D} > \frac{3}{4}$ compared to the calculated probability that $\bar{D} > \frac{3}{4}$ using the normal distribution.