W203 Lab2

Suhas Gupta

10/20/2018

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

Let F denote a fair coin and T a trick coin. Given that there are 99 fair coins & 1 trick coin in the bucket, we can write

$$P(F) = \frac{99}{100} \qquad \text{F : Event of selecting a fair coin}$$
$$P(T) = \frac{1}{100} \qquad \text{T: Event of selecting a trick coin}$$

(a)

H_k is the event that all heads occur in k flips (i.e., k heads in a row). From the theorem of conditional probability, we have:

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

Also,

$$P(T \cap H_k) = P(H_k \cap T) = P(H_k|T)P(T)$$

Conditional probability of getting k heads in a row given a trick coin:

$$P(H_k|T) = 1 \text{ (since the trick coin always comes up heads)}$$

Thus,

$$P(T \cap H_k) = 1 \cdot \frac{1}{100} = \frac{1}{100}$$

Also, we can write the conditional probability of getting k heads in a row given a fair coin as:

$$P(H_k|F) = \left(\frac{1}{2}\right)^k$$

Now, from the law of total probability we can write:

$$\begin{aligned} P(H_k) &= P(H_k|T)P(T) + P(H_k|F)P(F) \\ &= 1 \cdot \frac{1}{100} + \left(\frac{1}{2}\right)^k \cdot \frac{99}{100} \\ &= \frac{1}{100} + \frac{99}{100 \cdot 2^k} \\ &= \frac{1}{100}\left(1 + \frac{99}{2^k}\right) \end{aligned}$$

Thus the conditional probability that the coin is a trick coin given k heads in a row:

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

$$= \frac{\frac{1}{100}}{\frac{1}{100} \cdot \left(1 + \frac{99}{2^k}\right)}$$

$$\implies P(T|H_k) = \frac{1}{1 + \left(\frac{99}{2^k}\right)}$$

It is evident from the above expression that as k increases the conditional probability increases and converges to 1. This makes sense intuitively as the probability of the coin being a trick coin is larger if we get all heads in a row for a large number of trials.

(b)

For finding the value of k such that the above conditional probability is at least 99%, we solve the following inequality for k:

$$P(T|H_k) \geq \frac{99}{100}$$

$$\implies \frac{1}{1 + \left(\frac{99}{2^k}\right)} \geq \frac{99}{100}$$

$$\implies 1 + \left(\frac{99}{2^k}\right) \leq \frac{100}{99}$$

$$\implies \left(\frac{99}{2^k}\right) \leq \frac{1}{99}$$

$$\implies 2^k \geq 99^2$$

$$\implies k\log(2) \geq 2log(99)$$

$$\implies k \geq 2 \cdot \frac{log(99)}{log(2)}$$

$$\implies k \geq 13.26$$

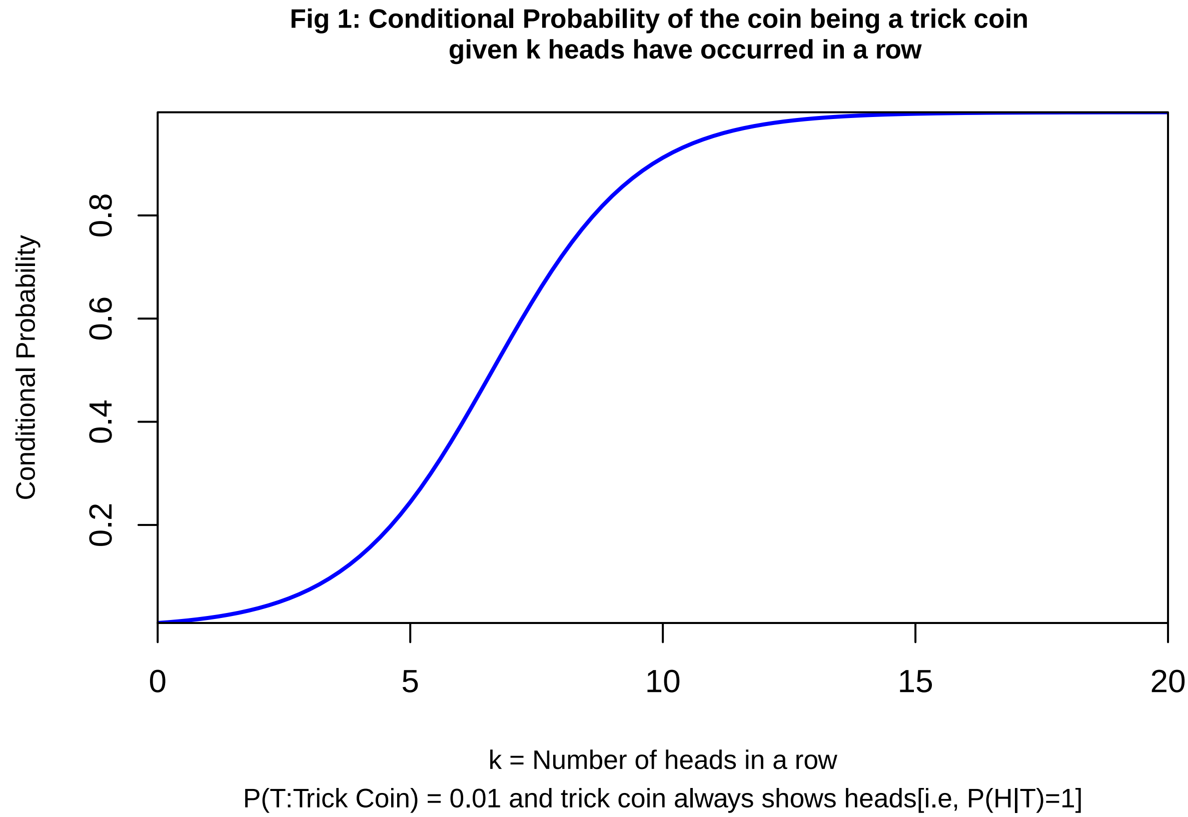
Since k is an integer (number of coin flips),

$$\boldsymbol{k_{min} = 14}$$

Thus the minimum number of heads in a row required is **14** for the conditional probability (of having a trick coin given k heads have occurred in a row) to be at least 99%.

Fig1. graphs this conditional probability as a function of k:

```
curve((1/((99/2^x)+1)),0,20,xaxs="i",yaxs="i",
      main = "Fig 1: Conditional Probability of the coin being a trick coin
      given k heads have occurred in a row",
      sub = "P(T:Trick Coin) = 0.01 and trick coin always shows heads[i.e, P(H|T)=1]",
      ylab = 'Conditional Probability', xlab = 'k = Number of heads in a row',
      cex.main=0.8, cex.sub=0.8,cex.lab=0.8,col="blue",lwd=2)
```



Question 2: Wise Investments

Given a random variable X has a binomial distribution with parameters n & p.

(a)

The probability mass function of X can be denoted by *B(x; n,p)* where n is the number of companies and p is the probability of success (i.e. company reaches unicorn status). The p.m.f is defined as the probability of selecting x successes and (n-x) failures from n outcomes. Here,

$$x \in 0,1,2 \text{ and } n = 2$$

Thus:

$$b(x; n, p) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x} & , \quad x \in (0, 1, 2) \\ 0 & , \quad otherwise \end{cases}$$

Since n = 2 & p = 0.75, we can write:

$$b(x; 2, 0.75) = \begin{cases} \binom{2}{x} \cdot (0.75)^x \cdot (0.25)^{2-x} & , \quad x \in (0, 1, 2) \\ 0 & , \quad otherwise \end{cases}$$

(b)

Cumulative probability function of X can be derived as:

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

$$F(0) = P(x = 0) = \binom{2}{0} (0.75)^0 \cdot (0.25)^2 = \frac{1}{16}$$

$$F(1) = F(0) + P(x=1) = \frac{1}{16} + \binom{2}{1}(0.75)^1 \cdot (0.25)^1 = \frac{7}{16}$$

$$F(2) = F(1) + P(x=2) = \frac{7}{16} + \binom{2}{2}(0.75)^2 \cdot (0.25)^0 = \frac{7}{16} + \frac{9}{16} = 1$$

The cumulative probability function of X can be written compactly as follows:

$$F(X) = \begin{cases} 1/16 & , \quad x = 0 \\ 7/16 & , \quad x \leq 1 \\ 1 & , \quad x \leq 2 \end{cases}$$

(c)

Expectation of X:

$$E(X) = \sum_{x=0}^{2} x \cdot p(x) = \sum_{x=1}^{2} x \cdot p(x)$$
$$= 1 \cdot b(1; 2, 0.75)|2 \cdot b(2; 2, 0.75)$$
$$= 1 \cdot \frac{6}{16} + 2 \cdot \frac{9}{16}$$
$$\implies E(X) = 1.5$$

(d)

Variance of X ,

$$V(X) = E[(X - \mu)^2] = E(X^2) - [E(X)]^2$$

Now,

$$E(X^2) = \sum_{x=0}^{2} x^2 \cdot p(x)$$
$$= 0 + 1^2 \cdot \frac{6}{16} + 2^2 \cdot \frac{9}{16} = \frac{42}{16} = 2.625$$

$$\implies V(X) = 2.625 - 1.5^2 = 0.375$$

Question 3:Relating Min and Max

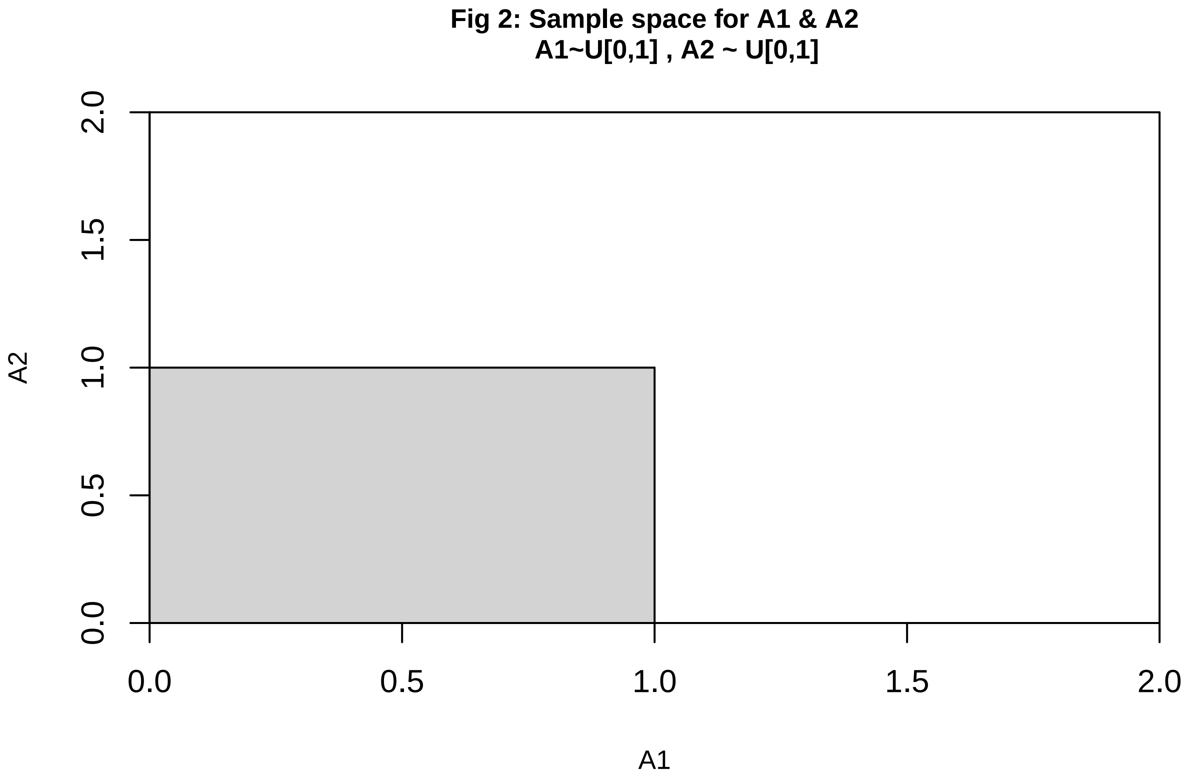
The joint probability function of X & Y is provided as:

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

(a)

Given, X=max{A1,A2} and Y= min{A1,A2} are functions of two uniformly distributed random variables A1~U[0,1] & A2~U[0,1]. We first draw the sample space of A1 and A2 as a square in the xy plane with each side of length 1 (shown in Fig2):

```
x0 = c(0,0.1, 1)
y0 = c(0,0.1,1)
curve(x*1,0,1, xlim=c(0,2), ylim=c(0,2),xaxs="i", yaxs="i", ylab="A2", xlab="A1",
      main="Fig 2: Sample space for A1 & A2
      A1~U[0,1] , A2 ~ U[0,1]", cex.main=0.8, cex.lab=0.8)
rect(0,0,1,1,col='lightgray')
```

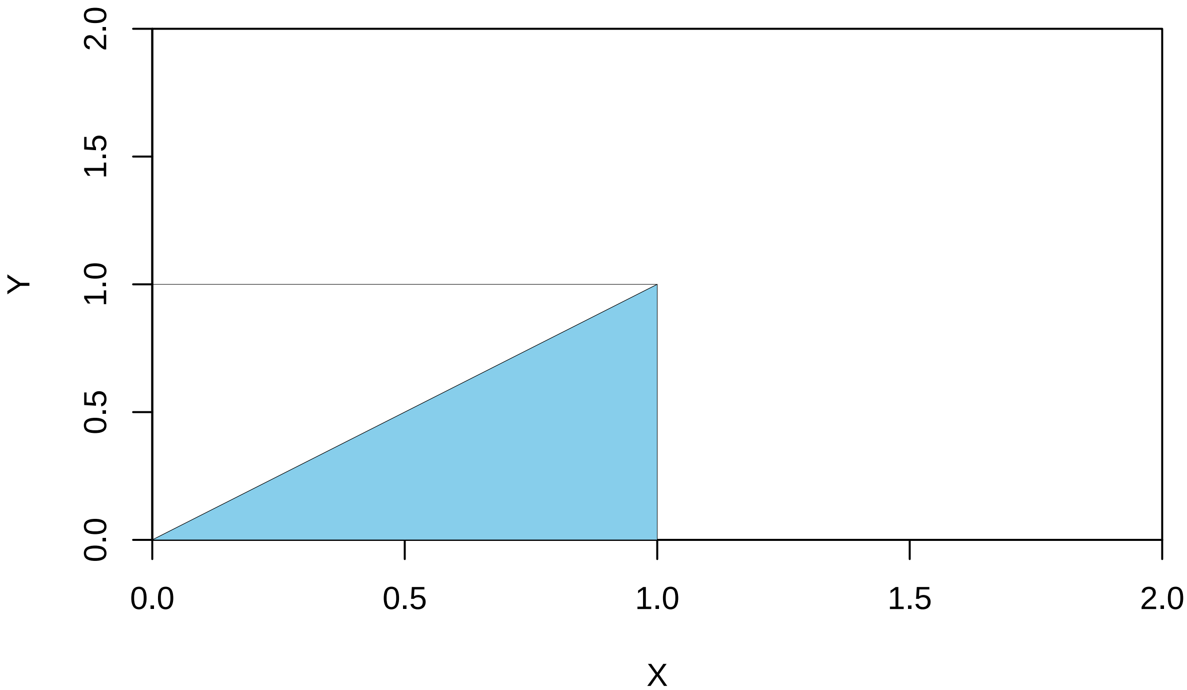


Since $X = max(A1, A2)$ and $Y = min(A1, A2)$, we realize that X & Y have the same support as A1,A2 and that Y < X. We thus draw the region of positive probability density for X & Y as the right half triangle across the sample space square diagonal (light blue shaded region in Fig3).

```
x0 <- c(0,1,1)

y0 <- c(0,1,0)
curve(x*1,0,1,  xlim=c(0,2), ylim=c(0,2),xaxs='i',yaxs='i',xlab="X", ylab="Y",
      main="Fig 3: Region of positive probability density of X & Y",
      sub="X=max(A1,A2) & Y=min(A1,A2) where A1 & A2 ~U(0,1)",
      cex.main=0.8, cex.lab=1, cex.sub=0.8,lwd=0.2)
rect(0,0,1,1,lwd=0.2)
polygon(x0,y0,col='skyblue',lwd=0.2)
```

Fig 3: Region of positive probability density of X & Y



X=max(A1,A2) & Y=min(A1,A2) where A1 & A2 ~U(0,1)

(b)

The marginal probability function of X, f(x) is given by:

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

$$f_X(x) = \int_{0}^{x} 2 \cdot dy = 2y\Big|_0^x \quad \text{for } 0 < x < 1$$

$$\implies f_X(x) = \begin{cases} 2x & , \quad 0 < x < 1 \\ 0 & , \quad otherwise \end{cases}$$

(c)

The unconditional expectation of X is given by:

$$E(X) = \int_{-\infty}^{\infty} x \cdot f(x)\ dx$$

$$= \int_0^1 x \cdot 2x\ dx = \int_0^1 2x^2\ dx = \frac{2}{3}x^3\Big|_0^1$$

$$\implies E(X) = \frac{2}{3} = 0.67$$

(d)

The probability density function of Y conditional on X, is given by:

$$f_{Y|X}(y|x) = \frac{f(x,y)}{f_X(x)} \quad \text{for} - \infty < y < \infty$$

$$f_{Y|X}(y|x) = \begin{cases} \frac{2}{2x} = \frac{1}{x} & , \quad 0 < y < x < 1 \\ 0 & , \quad otherwise \end{cases}$$

(e)

The conditional expectation of Y if given by :

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

$$= \int_0^x y \cdot \frac{1}{x}dy = \frac{1}{x}\int_0^x y\ dy = \frac{1}{2x}y^2\Big|_0^x$$

$$\implies E(Y|X) = \frac{x}{2}$$

(f)

Using the law of iterated expectations, we can write:

$$E(XY) = E_X[E(XY|X)] = E_X[XE(Y|X)]$$

The above algebraic simplification uses the fact that X is constant when computing expectation of X conditional on itself.

$$= E[X \cdot \frac{X}{2}] = E(\frac{X^2}{2}) = \frac{1}{2}E(X^2)$$

$$= \frac{1}{2}\int_{-\infty}^{\infty} x^2 \cdot f(x)dx = \frac{1}{2}\int_0^1 x^2 \cdot 2x\ dx = \int_0^1 x^3\ dx$$

$$= \frac{x^4}{4}\Big|_0^1 = \frac{1}{4} = 0.25$$

(g)

Co-variance of X & Y is given as:

$$cov(X,Y) = E(XY) - E(X)E(Y)$$

Using the law of iterated expectations we can write E(Y) as:

$$E(Y) = E_X[E(Y|X)] = \int_{-\infty}^{\infty} E(Y|X) \cdot f(x)\ dx$$

$$E(Y) = \int_0^1 \frac{x}{2} \cdot 2x\ dx = \int_0^1 x^2\ dx = \frac{1}{3}$$

Now we can compute co-variance:

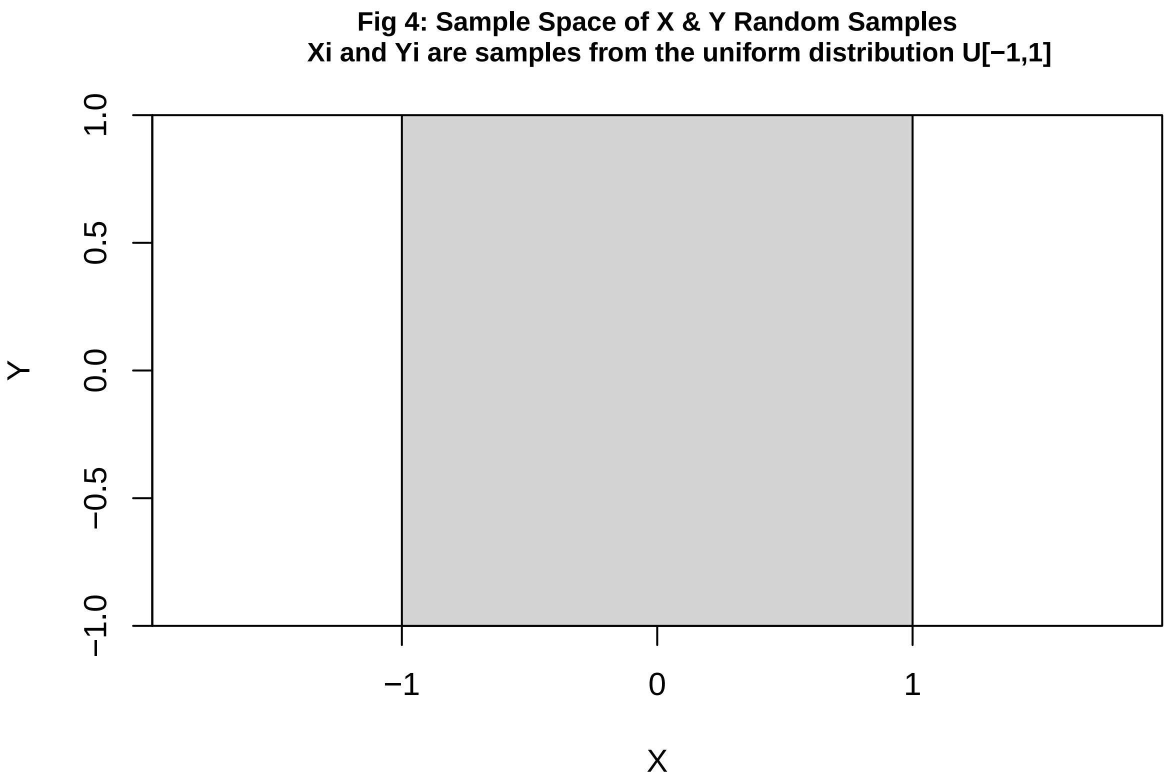
$$cov(X,Y) = E(XY) - E(X)E(Y) = \frac{1}{4} - \frac{2}{3}\cdot\frac{1}{3} = \frac{1}{4} - \frac{2}{9}$$

$$\implies cov(X,Y) = \frac{1}{36} = 0.0278$$

Question 4: Circles, Random Samples, and the Central Limit Theorem

The samples for random variables X & Y are uniformly distributed in [-1,1]. Thus, we can draw the sample space for X_i & Y_i as a square in xy plane with side length 2 (Fig 4):

```
x0 = c(-1,0.1,1)
y0 = c(-1,0.1,1)
curve(x*1,0,1, xlim=c(-1,1),ylim=c(-1,1), ylab="Y", xlab="X", xaxs="i",yaxs="i",
      main="Fig 4: Sample Space of X & Y Random Samples
      Xi and Yi are samples from the uniform distribution U[-1,1]",
      asp=1,cex.main=0.8, cex.sub=0.8,cex.lab=1)
rect(-1,-1,1,1,col='lightgray')
```

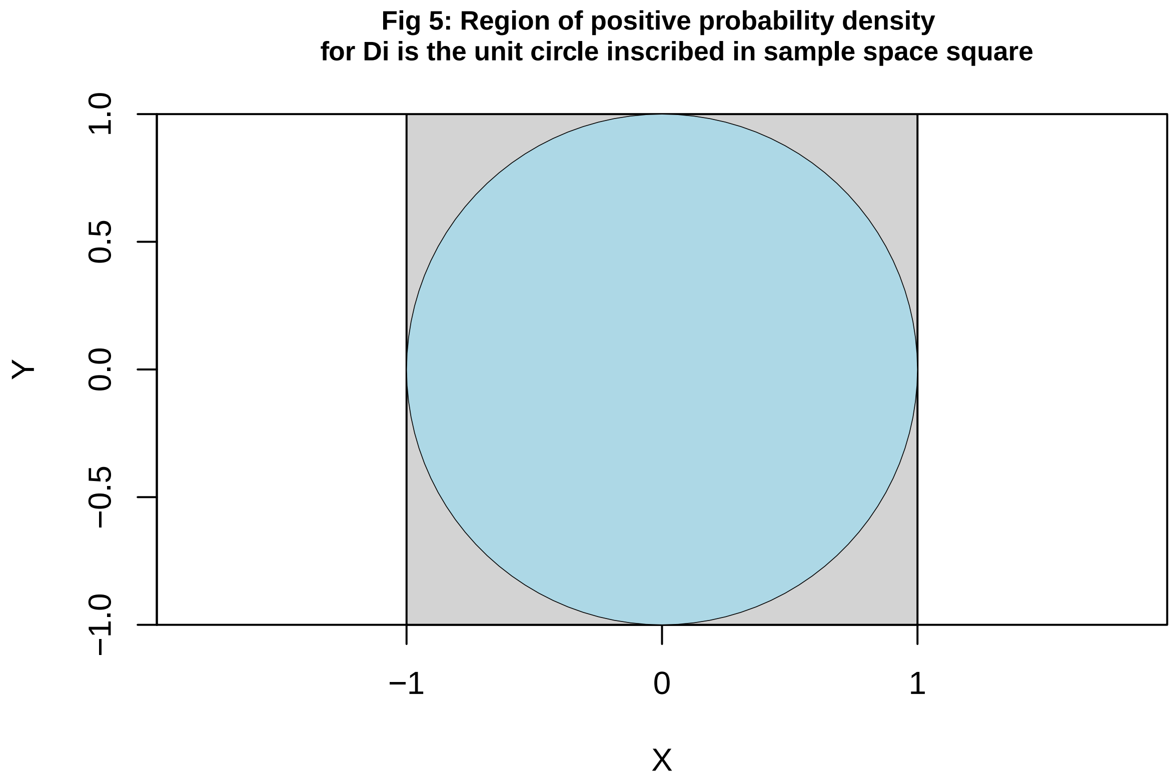


D_i is the Bernoulli random variable with successful outcome when $X_i^2 + Y_i^2 < 1$.

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

Using the definition of Bernoulli variable, we can define the probability of success as the probability of X_i, Y_i lying inside a unit circle. The region of positive probability density for D_i is thus a unit circle inscribed within the sample space square (blue shaded region in Fig 5)

```
x0 = c(-1,0.1,1)
y0 = c(-1,0.1,1)
curve(x*1,0,1, xlim=c(-1,1),ylim=c(-1,1), ylab="Y", xlab="X", xaxs="i",yaxs="i",
      main="Fig 5: Region of positive probability density
    for Di is the unit circle inscribed in sample space square",
      asp=1, cex.main=0.8, cex.lab=1)
rect(-1,-1,1,1,col='lightgray')
draw.circle(0,0,1,nv=100,border=NULL,col="lightblue",lty=1,density=NULL,
  angle=45,lwd=0.4)
```



Let p denote the probability of success for the Bernoulli random variable D_i. We can calculate the value of p as the ratio of area of unit circle (region of positive probability density) to the area of the square (sample space for X & Y) (see Fig 5):

$$p = \frac{\text{Area of Circle}}{\text{Area of sample space}}$$
$$\implies p = \frac{\pi \cdot (1)^2}{2 \cdot 2}$$
$$\implies p = \frac{\pi}{4}$$

(a)

We know that the expectation of a Bernoulli random variable with parameter p is equal to p. Thus,

$$E(D_i) = p = \frac{\pi}{4}$$

(b)

Variance of Bernoulli random variable with parameter p is equal to $p(1-p)$.

Thus, we have

$$\sigma^2_{D_i} = p(1-p)$$
$$\implies \sigma_{D_i} = \sqrt{p(1-p)}$$

Substituting the value of p, we get

$$\implies \sigma_{D_i} = \sqrt{\frac{\pi}{4}\left(1 - \frac{\pi}{4}\right)}$$

$$\implies \sigma_{D_i} = \frac{1}{4}\sqrt{\pi(4-\pi)}$$

(c)

The sample average can be written as :

$$\bar{D} = \frac{1}{n}\sum_{i=1}^{n} D_i \tag{1}$$

The sample variance is given by,

$$S^2 = Var\Big(\frac{1}{n}\sum_{i=1}^{n} D_i\Big) = \frac{1}{n^2}\Big[Var\Big(\sum_{i=1}^{n} D_i\Big)\Big]$$

For i.i.d variables the variance of sum is equal to the sum of variances. Thus above expression can be simplified as:

$$= \frac{1}{n^2}\cdot n\Big[Var(D_i)\Big] = \frac{np(1-p)}{n^2}$$

$$S^2 = \frac{p(1-p)}{n}$$

The standard error is the standard deviation of the sample mean.

$$S = \sqrt{\frac{p(1-p)}{n}}$$

Substituting the value of p calculated above in the expression for S, we get

$$S = \frac{1}{4}\sqrt{\frac{\pi(4-\pi)}{n}}$$

(d)

The sample average is given by the expectation,

$$\mu_{\bar{D}} = \mu = p = \frac{\pi}{4} = 0.7854 \tag{2}$$

The sample std. deviation (std. error),

$$\sigma_{\bar{D}} = S = \frac{1}{4}\sqrt{\frac{\pi(4-\pi)}{n}} = 0.041 \tag{3}$$

According to the central limit theorem (CLT) if n is sufficiently large, the sample mean will have a normal distribution that is centered at the mean of the population with std. deviation derived above. Since n=100 in our case, we can use the CLT to approximate the required probability:

$$P(\bar{D} \geq d) \approx P\Big(Z \geq \frac{d-\mu_{\bar{D}}}{\sigma_{\bar{D}}}\Big)$$

$$\implies P(\bar{D} \geq \frac{3}{4}) \approx P\Big(Z \geq \frac{0.75 - 0.7854}{0.041}\Big) = 1 - P\Big(Z \leq \frac{0.75 - 0.7854}{0.041}\Big) = 1 - P(Z \leq -0.86)$$
$$= 1 - \phi(-0.86)$$

Using the z-distribution tables we can calculate the area under a normal curve critical value derived above.

Thus,

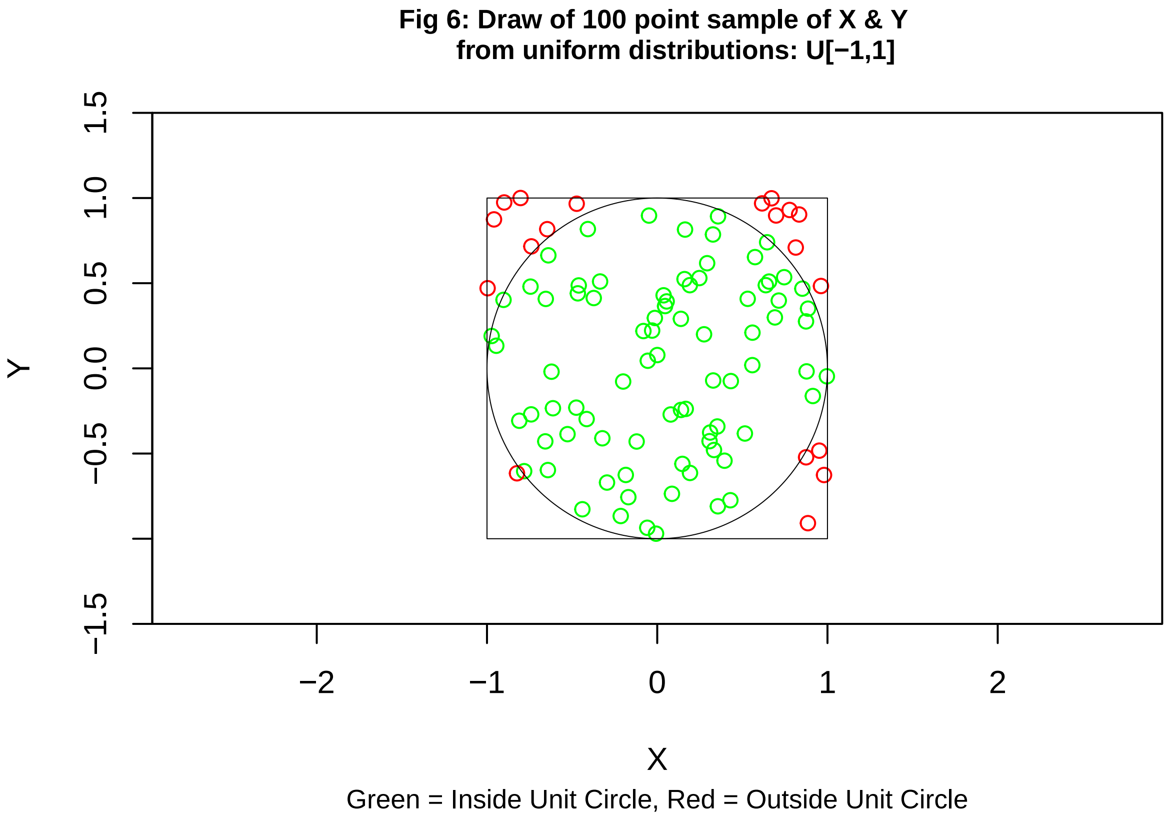
$$P(\bar{D} \geq \frac{3}{4}) = 1 - 0.1949 = 0.8051$$

(e)

```
# Draw random sample of X & Y from uniform distributions
n=100
X = runif(n,min=-1,max=1)
Y = runif(n,min=-1,max=1)
# Define the function to compute bernoulli random variable D
uniform_circ_dist = function(n,p){
    # Draw random sample of X & Y from uniform distributions
    X = runif(n,min=-1,max=1)
    Y = runif(n,min=-1,max=1)
    bernoulli_outcome = c()
    ## Compute the bernoulli variable D from X & Y values
    for (i in 1:n) {
        if ((X[i]^2+Y[i]^2)<1){
            bernoulli_outcome[i] = 1
        }
        else{
            bernoulli_outcome[i] = 0
        }
    }
    # return the vector of outcomes for bernoulli variable
    return(list(X=X,Y=Y,D=bernoulli_outcome))
}

n = 100
p = pi/4
result = uniform_circ_dist(n,p)
X = unlist(X)
Y = unlist(Y)
D = as.numeric(unlist(result[3]))
## Plot the X & Y data points
## and color them based on their position relative to the unit circle
dataColor=c()
dataColor[(X^2+Y^2)<1] = "green"
dataColor[(X^2+Y^2)>=1] = "red"
plot(X,Y, col=dataColor, asp=1, xaxs="i",yaxs="i",
     xlim =c(-1.5,1.5), ylim=c(-1.5,1.5),
     cex.main=0.8, cex.lab=1,cex.sub=0.8,
     main ="Fig 6: Draw of 100 point sample of X & Y
     from uniform distributions: U[-1,1]", xlab="X", ylab = "Y",
     sub = "Green = Inside Unit Circle, Red = Outside Unit Circle"
    )
# Superimpose a unit circle and sample space sqaure
draw.circle(0,0,1,nv=1000,border=NULL,lty=1,density=NULL,
```

```
  angle=45,lwd=0.5)
rect(-1,-1,1,1,lwd=0.5)
```



(f)

```
df = data.frame(p,mean(D), sprintf("%0.3f%%",abs(1-mean(D)/p)*100))
kable(df, booktabs = T,position = "center", align = "c",
      format= "pandoc",
      digits = 3,
      caption = "Comparison of simulated and algebraic sample averages of
      Bernoulli Di (Sample consists of 100 draws of Xi,Yi ~ U[-1,1])",
      col.names=(c("Simulated ","Algebraic","Percent Difference"))
      )
```

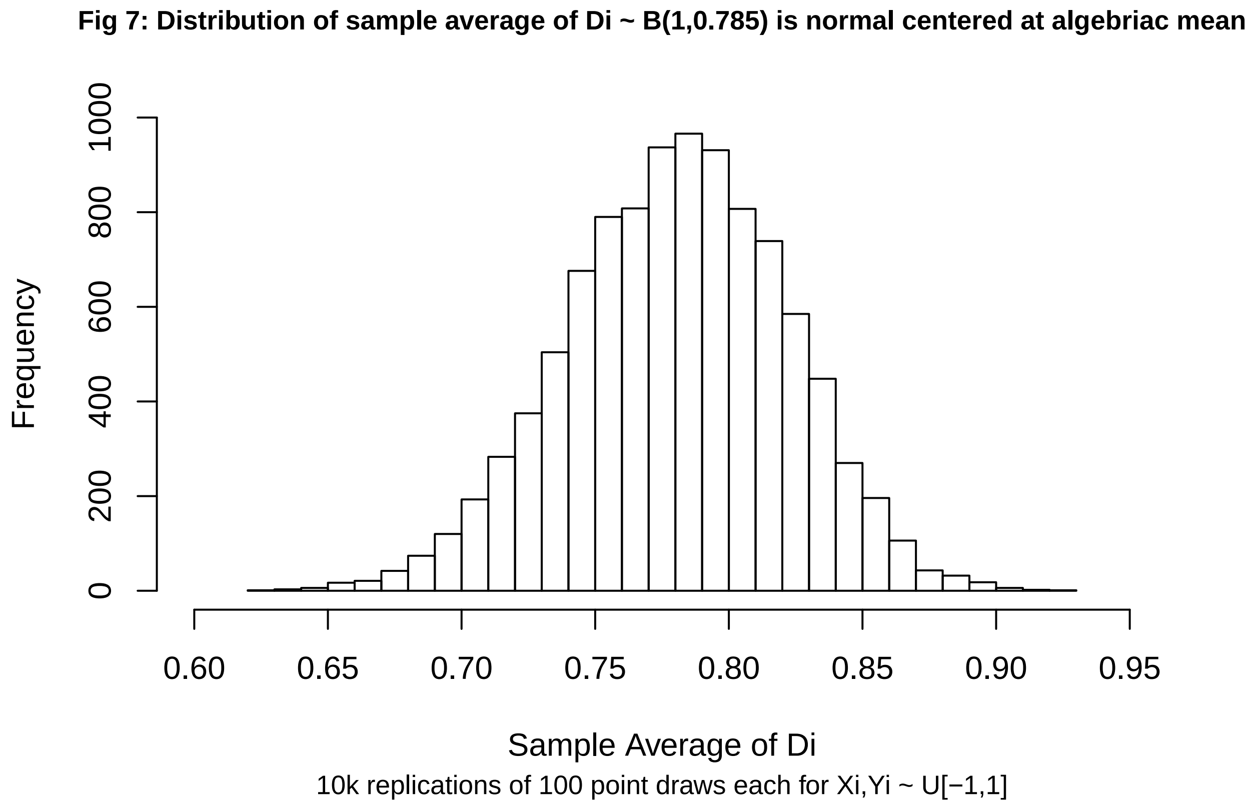
Table 1: Comparison of simulated and algebraic sample averages of Bernoulli Di (Sample consists of 100 draws of Xi,Yi ~ U[-1,1])

Simulated	Algebraic	Percent Difference
0.785	0.76	3.234%

The simulated sample average is within 1% of the algebraically computed sample average.

(g)

```r
n = 100
p = pi/4
reps = 10000
    X = runif(n,min=-1,max=1)
    Y = runif(n,min=-1,max=1)
mean_sample_averages = replicate(reps, mean(unlist(uniform_circ_dist(n,p)[3])))
hist(mean_sample_averages, breaks=25, xlim=c(0.60,0.95),ylim=c(0,1000), cex.main=0.8, cex.lab=1,cex.sub=
     main = "Fig 7: Distribution of sample average of Di ~ B(1,0.785) is normal centered at algebriac me
     sub = "10k replications of 100 point draws each for Xi,Yi ~ U[-1,1]",
     xlab = "Sample Average of Di")
```



The mean of the sample average is centered around the algebraic mean (=0.785) and has a normal distribution (Fig 7).

(h)

```r
# Theoretical variance and standard error of the sample average
sample_var = p*(1-p)/n
stderr = sqrt(sample_var)
df = data.frame(sd(mean_sample_averages),stderr,
                sprintf("%0.3f%%",100*abs(1- sd(mean_sample_averages)/stderr)))
kable(df, format= "pandoc", booktable = T, longtable = T,
      align="c", position="center",digits = 5,cex.main = 0.8,
      caption="Comparison of simulated and algebraic standard errors for
    Di with 10k repititions of 100 point samples of Xi,Yi ~ U[-1,1].
    (Di ~ B(1,0.785) with success defined as the event
    when Xi,Yi lie inside a unit circle)",
      col.names=(c("Simulated Std. Error","Algebraic Std. Error","Percent Difference")))
```

Table 2: Comparison of simulated and algebraic standard errors for Di with 10k repititions of 100 point samples of Xi,Yi ~ U[-1,1]. (Di ~ B(1,0.785) with success defined as the event when Xi,Yi lie inside a unit circle)

Simulated Std. Error	Algebraic Std. Error	Percent Difference
0.04126	0.04105	0.509%

The simulated value of standard error is within 1% of the algebraically calculated standard error for the statistic.

(i)

```
a = length(which(mean_sample_averages >= 0.75))
sim_prob = a/reps
df <- data.frame(sim_prob,
                 1-pnorm(-0.86), sprintf("%.3f%%",100*abs(1-sim_prob/(1-pnorm(-0.86)))))
kable(df, booktabs = T, position = "center", align = "c",
      format= "pandoc",
      digits = 3,
      caption = "Comparison of simulated and algebraic probabilities
    for sample average of Di being greater than 0.75 (Di ~ B(1,0.785;
    10k repititions of 100 point sample draws of Xi,Yi were performed; Xi,Yi ~ U[-1,1])",
      col.names=(c("Simulated Probability","Algebraic Probability","Percent Difference"))
      )
```

Table 3: Comparison of simulated and algebraic probabilities for sample average of Di being greater than 0.75 (Di ~ B(1,0.785; 10k repititions of 100 point sample draws of Xi,Yi were performed; Xi,Yi ~ U[-1,1])

Simulated Probability	Algebraic Probability	Percent Difference
0.836	0.805	3.850%

The simulated value is within 5% of the algebraic sample average (=0.805) that we calculated using the CLT approximation for the sample statistic.