W203 | Fall 2018 | WED 4 P.M.

LAB 2 Probability Theory | Atit Wongnophadol

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

Let $T$ = the event of selecting the trick coin.
$F$ = " " fair coin.
$H_k$ = the event that the coin comes up head all k times.

And $P(T) = 0.01$ , $P(F) = 0.99$
$P(H_k|T) = 1$ , $P(H_k|F) = (0.5)^k$

(a.)
$$P(T|H_k) = \frac{P(T \cap H_k)}{P(H_k)}$$
$$= \frac{P(H_k|T) \cdot P(T)}{P(H_k|T) \cdot P(T) + P(H_k|F) \cdot P(F)}$$
$$= \frac{1 \cdot (0.01)}{1 \cdot (0.01) + (0.5)^k (0.99)}$$
$$= \frac{1}{1 + (0.5)^k 99}$$

(b.) Find k such that $P(T|H_k) > 0.99$

$$\frac{1}{1 + (0.5)^k 99} > 0.99$$
$$1 > 0.99 + (0.5)^k (99)(0.99)$$
$$0.01 > 0.5^k (99)(0.99)$$
$$\frac{0.01}{99(0.99)} > 0.5^k$$
$$99^{-2} > 0.5^k$$
$$-2 \ln 99 > k \ln 0.5$$
$$\therefore k > -\frac{2 \ln 99}{\ln 0.5} \approx 13.2587$$

Since k must be Integer, $k \geq 14$ ensures that $P(T|H_k) > 0.99$.

## 2.) Wise Investments

a. The p.m.f. of $X$ is

$$f(x) = \begin{cases} \binom{2}{x}\left(\frac{3}{4}\right)^{x}\left(\frac{1}{4}\right)^{2-x} & ; x \in \{0, 1, 2\} \\ 0 & ; \text{otherwise} \end{cases}$$

b. The c.d.f. of $X$

$$f(0) = \binom{2}{0}\left(\frac{3}{4}\right)^{0}\left(\frac{1}{4}\right)^{2} = \frac{1}{16}$$

$$f(1) = \binom{2}{1}\left(\frac{3}{4}\right)^{1}\left(\frac{1}{4}\right)^{1} = \frac{6}{16}$$

$$f(2) = \binom{2}{2}\left(\frac{3}{4}\right)^{2}\left(\frac{1}{4}\right)^{0} = \frac{9}{16}$$

$$F(0) = f(0) = \frac{1}{16}$$

$$F(1) = f(0) + f(1) = \frac{7}{16}$$

$$F(2) = f(0) + f(1) + f(2) = 1$$

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

c. $$E(X) = \sum_{i=0}^{2} x_i f(x_i)$$

$$= (0)\left(\frac{1}{16}\right) + (1)\left(\frac{6}{16}\right) + (2)\left(\frac{9}{16}\right)$$

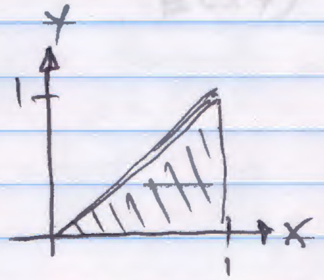
$$= \frac{3}{2} = 1.5$$

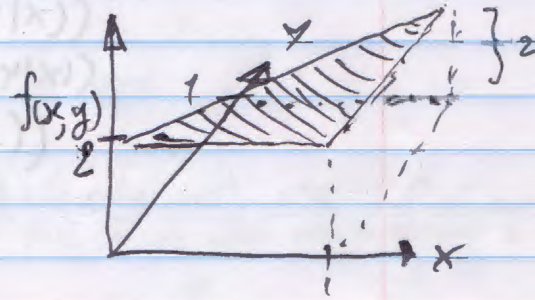
D. $$Var(X) = E(X^2) - [E(X)]^2$$

$$= \left[(0^2)\left(\frac{1}{16}\right) + (1^2)\left(\frac{6}{16}\right) + (2^2)\left(\frac{9}{16}\right)\right] - \left(\frac{3}{2}\right)^2 = \frac{21}{8} - \frac{18}{8} = \frac{3}{8}$$

③ Relating Min and Max

a.




b. $$f_X(x) = \int_{-\infty}^{\infty} f(x,y)\,dy$$

$$= \int_0^x 2\,dy$$

$$= 2|y|_0^x \quad \cancel{= 2x}$$

$$= 2x$$

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

$$= 2\int_{-\infty}^{\infty} x^2\,dx = 2\int_0^1 x^2\,dx$$

$$= 2\cdot\left|\frac{x^3}{3}\right|_0^1$$

$$= \frac{2}{3}.$$

d. $$f_{Y|X}(y|x) = \frac{f_{X,Y}(x,y)}{f_X(x)}$$

$$= \frac{2}{2x}$$

$$= \frac{1}{x}$$

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

$$= \int_{-\infty}^{\infty} \frac{y}{x}\,dy = \frac{1}{x}\int_0^1 y\,dy$$

$$= \frac{1}{x}\left|\frac{y^2}{2}\right|_0^1 = \frac{1}{2x}$$

f.

$$E(XY) = E(E(XY|X))$$
$$= E(X\,E(Y|X))$$
$$= E\left(X\left(\tfrac{1}{2X}\right)\right)$$
$$= \frac{1}{2}$$

g. Derive $cov(X,Y) = E(XY) - E(X)E(Y)$

~~First~~ we already know $E(XY)$ and $E(X)$.

Lets ~~first~~ find $E(Y) = E(E(Y|X))$

$$= E\left(\frac{1}{2X}\right) = \frac{1}{2}E\left(\frac{1}{X}\right)$$

$$= \frac{1}{2}\int_0^1 \frac{1}{x}\,\cancel{dx} \qquad = \frac{1}{2}\int_0^1 \frac{1}{x}(2x)\,dx$$

$$\cancel{= \frac{1}{2}\left|\ln x\right|_0^1} \qquad = \frac{1}{2}(2)\left|x\right|_0^1$$

$$= 1$$

$$cov(X,Y) = E(XY) - E(X)E(Y)$$
$$= \frac{1}{2} - \left(\frac{2}{3}\right)(1)$$
$$= -\frac{1}{6}$$

## ④ Circles

a. $D \sim$ Bernoulli with $p(1) = p$ and $p(0) = 1-p$

where $p = \dfrac{\text{Area of the circle with radius of 1}}{\text{Area of the square with the length of 2}}$

$$= \frac{\pi(1)^2}{(2)^2} = \frac{\pi}{4}$$

$$E(D) = 0 \cdot p(0) + (1) \cdot p(1)$$
$$= 0 \cdot (1-p) + p$$
$$\therefore E(D) = \frac{\pi}{4}$$

b. $Var(D) = p \cdot (1-p)$ following the Bernoulli Distribution

$$= \frac{\pi}{4}\left(1 - \frac{\pi}{4}\right)$$

$$\text{S.D. of } D \; (\sigma) = \sqrt{Var(D)}$$
$$\therefore \sigma = \frac{1}{4}\sqrt{4\pi - \pi^2} \sim 0.4105$$

c. $\sigma_{\bar{D}} = \sigma / \sqrt{n}$

$$= \frac{1}{4\sqrt{n}}\sqrt{4\pi - \pi^2} \sim \frac{0.4105}{\sqrt{n}}$$

d. $\mu_{\bar{D}} = \dfrac{\pi}{4}$

$$P(\bar{D} > \tfrac{3}{4}) = P\left(Z > \frac{\frac{3}{4} - \frac{\pi}{4}}{0.4105/\sqrt{100}}\right) = P(Z > -0.8622)$$
$$= 1 - \Phi(-0.8622) = 0.8057$$

The C.L.T. states that if $n$ is sufficiently large, the random sample $X$ from a given distribution approaces normal distribution with $\bar{x}$ as the mean and $\sigma_{\bar{x}}$ as the standard deviation $\sim N(\bar{x}, \sigma_{\bar{x}})$

In this problem, $D \sim$ Bernoulli distributed. As $n$ of 100 is large enough we can use the C.L.T. to approximate $\bar{D} \sim N(\mu_{\bar{D}}, \sigma_{\bar{D}})$.

# Lab 2: Probability Theory

W203: Statistics for Data Science

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

**e. Now let $n = 100$. Use R to simulate a draw for $X_1, X_2, ..., X_n$ and $Y_1, Y_2, ..., Y_n$. Calculate the resulting values for $D_1, D_2, ...D_n$. Create a plot to visualize your draws, with $X$ on one axis and $Y$ on the other. We suggest using a command like the following to assign a different color to each point, based on whether it falls inside the unit circle or outside it. Note that we pass $d+1$ instead of $d$ into the color argument because 0 corresponds to the color white.**

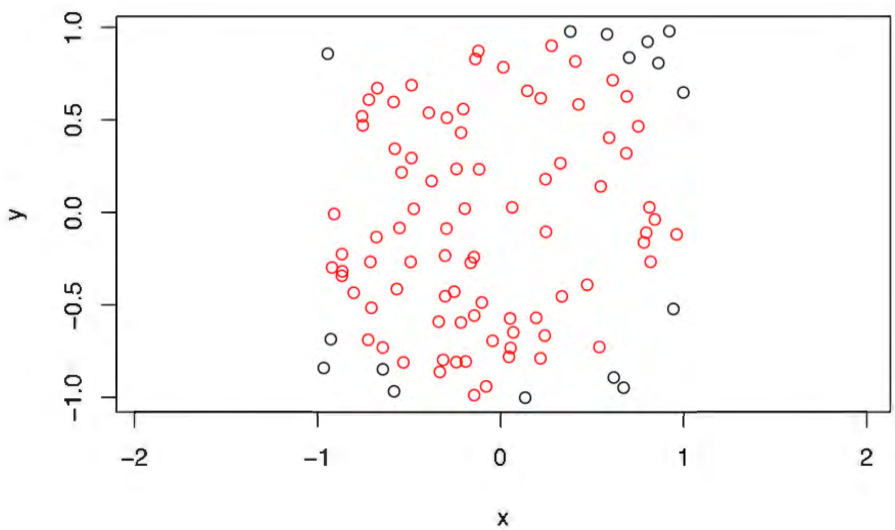
```r
# (Atit) Answer to question 4e:

# Set the number of simulation
n <- 100


# Generate random draws from a uniform distribution with range [-1,1]
# for both the r.v. X and the r.v. Y.
x <- runif(n, min=-1, max=1)
y <- runif(n, min=-1, max=1)


# Calculate the r.v. D.
d <- ifelse((x^2 + y^2) < 1, 1, 0)


# Plot X, Y, and D
plot(x,y, col=d+1, asp=1)
```




f. What value do you get for the sample average, $\bar{D}$? How does it compare to your answer for part a?

(Atit) Answer to question 4f:

The sample average, $\bar{D}$ is:

```
# Calculate mean of D
mean(d)
```

```
## [1] 0.84
```

The expected value of D from 4a is pi/4:

```
# Calculate mean of D
pi/4
```

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

The value is a bit different.

g. Now use R to replicate the previous experiment 10,000 times, generating a sample average of the $D_i$ each time. Plot a histogram of the sample averages.

```
##########################################################################
# Setup a function to generate a vector of D

simulate_d = function(n){
```

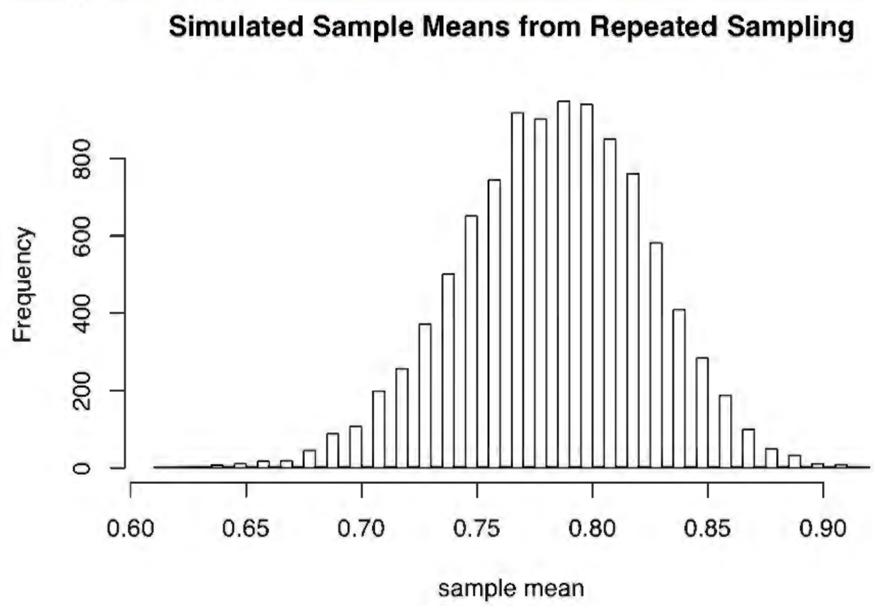
```r
# Generate random draws from a uniform distribution with range [-1,1]
# for both the r.v. X and the r.v. Y.
x <- runif(n, min=-1, max=1)
y <- runif(n, min=-1, max=1)

# Calculate the r.v. D.
sim_d <- ifelse((x^2 + y^2) < 1, 1, 0)

return(sim_d)
}

rep_d <- replicate(10000,mean(simulate_d(n)))

# Plot the histogram of the simulated sample means
hist(rep_d, breaks = 50, main = "Simulated Sample Means from Repeated Sampling",
     xlab = "sample mean")
```




h. Compute the standard deviation of your sample averages to see if it's close to the value you expect from part c.

**(Atit) Answer to question 4h:**

```r
# Compute standard deviation of the sample averages:
sd(rep_d)
```

```
## [1] 0.04125763
```

**This value is close the what is computed in 4c which was ~ 0.4105 / sqrt(100) = 0.04105.**

### i. Compute the fraction of your sample averages that are larger that $3/4$ to see if it's close to the value you expect from part d.

**(Atit) Answer to question 4i:**

```
# Compute the fraction of thhe sample averages that are larger than 3/4:
sum(rep_d[rep_d > 0.75])/sum(rep_d)
```

```
## [1] 0.7887416
```

**This value is close the what is computed in 4d which was 0.8057.**