Lab 2: Probability Theory

w203: statistics for Data Science

Adam Yang, Section 4.

1a. T = event you select a trick coin

H_K = event where all K flips are heads.

we know
$$\begin{cases} P(T) = 0.01 = \frac{1}{100} & \text{: We have 100 coins, 1 is a trick coin} \\ P(H_K \mid T) = 1 & \text{: Trick coin can only come up with heads} \\ P(H_K \mid T') = (0.5)^K & \text{: Given that we selected a fair coin,} \\ & \text{probability of getting only heads is } (0.5)^K \end{cases}$$

Solve $P(T \mid H_K) = \dfrac{P(T \wedge H_K)}{P(H_K)}$

$$P(T \wedge H_K) = P(H_K \wedge T) = P(H_K \mid T)\,P(T) = (1)(0.01) = 0.01$$

$$\begin{aligned} P(H_K) &= P(H_K \wedge T) + P(H_K \wedge T') \\ &= P(H_K \wedge T) + P(H_K \mid T')\,P(T') \\ &= 0.01 + (0.99)(0.5)^K \end{aligned}$$

$$\boxed{\therefore\ P(T \mid H_K) = \frac{0.01}{0.01 + (0.99)(0.5)^K}}$$

1b) How many flips to get $P(T|H_K) > 0.99$?

$$\frac{0.01}{0.01 + (0.99)(0.5)^K} > 0.99$$

$$\frac{0.01}{0.99} > 0.01 + (0.99)(0.5)^K$$

$$\frac{\left(\frac{0.01}{0.99} - 0.01\right)}{0.99} > 0.5^K$$

$$\ln\left(\frac{\frac{0.01}{0.99} - 0.01}{0.99}\right) > K \ln(0.5)$$

$$\frac{\ln\left(\frac{\frac{0.01}{0.99} - 0.01}{0.99}\right)}{\ln(0.5)} < K$$

$\ln(0.5)$ is negative, flip signs when dividing by a negative number

$$\therefore K > \frac{\ln\left(\frac{\frac{0.01}{0.99} - 0.01}{0.99}\right)}{\ln(0.5)} = 13.2587$$

$K > 13.2587$

$\boxed{\text{You need at least 14 heads in a row}}$

2a. Let S_1 = event where company 1 becomes a unicorn

Let S_2 = event where company 2 becomes a unicorn,

$P(S_1) = P(S_2) = \frac{3}{4}$ $\quad$ S_i and S_2 are independent.

Let x = total number of companies that become unicorns.

Possible outcomes, $x = 0 \quad (S_1' \cap S_2')$

$x = 1 \quad (S_1 \cap S_2') \cup (S_1' \cap S_2)$

$x = 2 \quad (S_1 \cap S_2)$

for $x=0$, $\left(\frac{1}{4}\right)\left(\frac{1}{4}\right) = \frac{1}{16}$

$x=1$, $\left(\frac{3}{4}\right)\left(\frac{1}{4}\right) + \left(\frac{1}{4}\right)\left(\frac{3}{4}\right) = \frac{6}{16}$

$x=2$, $\left(\frac{3}{4}\right)\left(\frac{3}{4}\right) = \frac{9}{16}$

$$\boxed{f(x) = \begin{cases} \frac{1}{16}, & x=0 \\ \frac{6}{16}, & x=1 \\ \frac{9}{16}, & x=2 \\ 0, & \text{otherwise} \end{cases}}$$

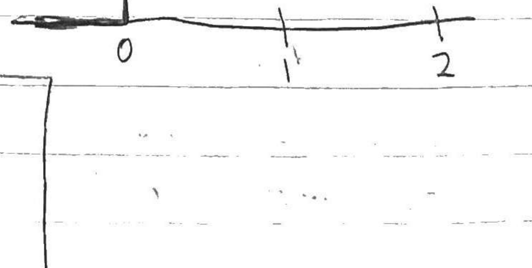
2b.

$0, \quad x<0$

$\frac{1}{16}, \quad 0 \le x < 1$

$\frac{1}{16} + \frac{6}{16}, \quad 1 \le x < 2$

$\frac{1}{16} + \frac{6}{16} + \frac{9}{16}, \quad x \ge 2$



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

2c) $E(x) = \sum_{i=1}^{2} x_i f(x_i) = \frac{1}{16}(0) + \left(\frac{6}{16}\right)(1) + \left(\frac{9}{16}\right)(2)$

$E(x) = 1.5$

2d) $Var(x) = E(x^2) - [E(x)]^2$

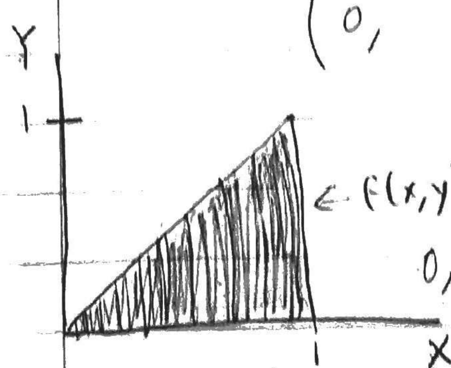
$E(x^2) = \sum_{i=1}^{2} x_i^2 f(x_i) = (0)^2\left(\frac{1}{16}\right) + (1)^2\left(\frac{6}{16}\right) + (2)^2\left(\frac{9}{16}\right)$

$= \frac{6}{16} + \frac{9(4)}{16} = \frac{42}{16} = 2.625$

$E(x)^2 = (1.5)^2 = \frac{9}{4} = 2.25$

$Var(x) = 2.625 - 2.25 = \underline{0.375}$

3a) $f(x,y) = \begin{cases} 2, & 0<y<x<1 \\ 0, & \text{otherwise.} \end{cases}$



3b) $f_x(x) = \int_{y=0}^{x} f_{x,y}(x,y)\,dy = \int_0^x 2\,dy = \Big[2y\Big]_0^x = 2x$

$\boxed{f_x(x) = \begin{cases} 2x, & 0<x<1 \\ 0, & \text{otherwise} \end{cases}}$

3c) $E(x) = \int_{-\infty}^{\infty} x f(x)\,dx = \int_0^1 2x^2\,dx = \left[\frac{2}{3}x^3\right]_0^1 = \frac{2}{3}$

$$\boxed{E(x) = \frac{2}{3}}$$

3d) $f_{Y|X}(y|x) = \frac{f_{X,Y}(x,y)}{f_X(x)} = \frac{2}{2x} = \frac{1}{x}$

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

3e) $E(Y|X) = \int_{-\infty}^{\infty} y\, f_{Y|X}(y|x)\,dy = \int_0^x \frac{y}{x}\,dy = \left[\frac{y^2}{2x}\right]_0^x = \frac{x^2}{2x}$

$$\boxed{E(Y|X=x) = \frac{x}{2} \quad \text{for } 0<x<1}$$

$E(Y|X=E(x)) = \frac{2}{3}\cdot\frac{1}{2} = \frac{1}{3}$

3f) $E(XY) = E(E(YX|X)) = E(xE(Y|X)) = E\left(\frac{x^2}{2}\right)$

$E\left(\frac{x^2}{2}\right) = \int_{-\infty}^{\infty} g(x) f_x(x)\,dx = \int_0^1 \frac{x^2}{2}\cdot 2x\,dx = \int x^3\,dx$

$= \left[\frac{1}{4}x^4\right]_0^1 \qquad \boxed{E(XY) = \frac{1}{4}}$

3g) $Cov(x,Y) = E(XY) - E(x)E(Y)$

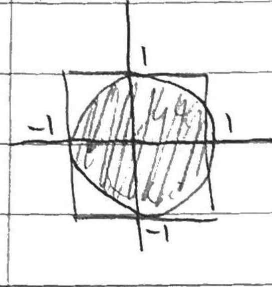
$E(Y) = E(E(Y|X)) = E\left(\frac{x}{2}\right) = \frac{1}{2}E(x) = \frac{1}{2}\cdot\frac{2}{3}$

$E(Y) = \frac{1}{3}$

$Cov(X,Y) = \frac{1}{4} - \left(\frac{2}{3}\right)\left(\frac{1}{3}\right) = \frac{1}{36} \qquad \boxed{Cov(X,Y) = \frac{1}{36}}$

4a) $D_i = \begin{cases} 1, & X_i^2 + Y_i^2 < 1 \\ 0, & \text{otherwise} \end{cases}$

Each D_i is a Bernoulli variable and are i.d.d.



← ratio of these areas is the probability that (X_i, Y_i) falls in the circle.

$$= \frac{\pi r^2}{2 \cdot 2} = \frac{\pi}{4}$$

$$f(d) = \begin{cases} \frac{\pi}{4}, & d=1 \\ 1-\frac{\pi}{4}, & d=0 \end{cases}$$

$$E(D_i) = \sum_{i=0}^{1} d_i f(d_i)$$
$$= \left(\frac{\pi}{4}\right)(1) + \left(1-\frac{\pi}{4}\right)(0) = \frac{\pi}{4}$$

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

4b) $Var(D_i) = E(D_i^2) - [E(D_i)]^2$

$$E(D_i^2) = \sum_{i=0}^{1} d^2 f(d) = (1)^2\left(\frac{\pi}{4}\right) + (0)^2\left(1-\frac{\pi}{4}\right) = \frac{\pi}{4}$$

$$\boxed{\begin{aligned} Var(D_i) &= \frac{\pi}{4} - \left(\frac{\pi}{4}\right)^2 \approx 0.1685 \\ \sigma &= \sqrt{\frac{\pi}{4} - \left(\frac{\pi}{4}\right)^2} \approx 0.4105 \end{aligned}}$$

4c) $E(D_i)$ and $Var(D_i)$ are know,

Central limit theorem says that is n is large,

$$\boxed{\sigma_{\bar{D}} = \frac{\sigma}{\sqrt{n}} = \frac{0.4105}{\sqrt{n}}}$$