Data Science Assignment

Ques1

(a) For $k = 6$.

First face value (FV)	Second face (remaining value)	[For remaining fa]	Probability first face
1	1, 2, 3, 4, 5, 6 →	$\lceil \frac{6}{2} \rceil$	$\times \frac{6}{6}$
2	2, 4, 6 →	$\lceil \frac{6}{2} \rceil = 3$	$\times \frac{3}{6}$
3	3, 6	$\lceil \frac{6}{3} \rceil = 2$	$\times \frac{2}{6}$
4	4	$\lceil \frac{6}{4} \rceil = 2$	$\times \frac{1}{6}$
5	5	$\lceil \frac{6}{5} \rceil = 2$	$\times \frac{1}{6}$
6	6	$\lceil \frac{6}{6} \rceil = 1$	$\times \frac{1}{6}$

Hence, for left side the probability of selecting each eliminate $\frac{1}{k} = \frac{1}{6}$ $(k=6)$

Probability of two independent event A, B is

$P(A \cap B) = P(A) \times P(B)$

$$= \frac{1}{6} \times \left[\frac{6}{6}\right] + \frac{1}{6} \times \frac{3}{6} + \frac{1}{6} \times \frac{2}{6} + \frac{1}{6} \times \frac{1}{6} + \frac{1}{6} \times \frac{1}{6} + \frac{1}{6} \times \frac{1}{6}$$

For FV=1 FV=2 FV=3 FV=4 FV=5 FV=6

$$= \frac{1}{6} \times \frac{6}{6} + \frac{1}{6} \times \frac{3}{6} + \frac{1}{6} \times \frac{2}{6} + \frac{1}{6} \times \frac{1}{6} + \frac{1}{6} \times \frac{1}{6}$$

$$= \frac{1}{6} + \frac{1}{6} \times \frac{1}{2} + \frac{1}{6} \times \frac{1}{3} + \frac{1}{6} \times \frac{1}{6} + \frac{1}{6} \times \frac{1}{6}$$

$$= \frac{1}{6} + \frac{1}{6} + \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{1 \times 5}{6} = \boxed{\frac{5}{6}}$$

Ans

Ques1(b) k is divisible by 1 to $\frac{k}{2}$ the integers.

For first face value ranging from 1 ... k

for first face is 1 → $\frac{k}{1}$ value can take excluding first.

$f.v = 2 \rightarrow \lfloor \frac{k}{2} \rfloor$ values $= \frac{k}{2}$

$f.v = 3 \rightarrow \lfloor \frac{k}{3} \rfloor = \frac{k}{3}$

$f.v = \frac{k}{2} \quad \lfloor \frac{k}{\frac{k}{2}} \rfloor = \frac{k}{\frac{k}{2}}$

$f.v. \rightarrow \frac{k}{\text{half of the value}} \quad \frac{k}{\frac{k}{2}} = 1$

Because the value k is not divisible by numbers more than k/2 to k. Only it ca be divided by the number itself therefore for numbers more than k/2 only value 1 since they are more than half of k

Hence, total probability is

first eg values all distributions

$$\frac{1}{k} \times \left[\frac{k}{k}\right] + \frac{1}{k} \times \left[\frac{k}{2}\right] + \frac{1}{k} \times \left[\frac{k}{3}\right] + \frac{1}{k} \times \left[\frac{k}{4}\right] + \frac{1}{k} \times \left[\frac{k}{\frac{k}{2}}\right]$$

k — half of k values

$$\int \frac{1}{x} = \frac{1}{1} + \frac{1}{2} + \cdots + \frac{1}{n} + \frac{1}{n+1}$$

$$\int_k^k \frac{1}{x}\,dx = \log x \Big|_k^k = \log k$$

$$\frac{\left[k + \frac{k}{2} - - - + \frac{k}{k/2}\right] + \frac{k}{2}}{k \times k}$$

$$= \frac{k\left[1 + \frac{1}{2} + \frac{1}{3} - + \frac{2}{k}\right] + \frac{k}{2}}{k^2}$$

$$= \frac{\left[1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{2}{k}\right] + \frac{1}{2}}{k} = \frac{\frac{1}{2} + \text{Harmonic sum}(k/2)}{k}$$

$$= \frac{\frac{1}{2} + \log k}{k}$$

Ques3 50% of email [at most]

$p_0 \leq .5$ [at most email spam]

55 spam emails out of n.

$$\frac{p - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} \leq 1.645$$

0.05

1.645

$p = \frac{55}{n}$

will not reject the hypothesis here.

$$\frac{\frac{55}{n} - (0.5)}{\sqrt{\frac{0.5(1-0.5)}{n}}} \leq 1.645$$

$$\frac{55 - 0.5n}{\sqrt{n}\,(0.5)} \leq 1.645$$

$$2(55 - 0.5n) \leq 1.645\sqrt{n}$$

$$n + 1.645\sqrt{n} \geq 110$$

$$(\sqrt{n} - 9.69779)(\sqrt{n} + 11.3478) \geq 0$$

$$\sqrt{n} \geq 9.69779$$

this has to be not more than zero equal to 0.

Since this is always +ve

$$n \geq 94.0471$$

$$\boxed{n = 95}$$ Ans

Ques 3b

H_0 : Proportion $\frac{2}{3}$

H_1 : Pop $< \frac{2}{3}$

$$3\left(\frac{29}{x} - \frac{1}{3}\right) < -1.645$$

7 cards → dice 28

$$\frac{\sqrt{2}}{\sqrt{x}}$$

$$\frac{(3 \times 29 - 1)}{x} < -1.645$$

$$\frac{(3 \times 29 - x)}{\sqrt{x} \times \sqrt{2}} < -1.645$$

$$3 \times 29 - x < -2.3264 \sqrt{2}$$

$$x - 2.3264 \times 1.\sqrt{2} + (3 \times 29) > 0$$

$$y^2 - 2.3264 \times y - 87 > 0$$

$$(y - 10.4012)(y + 8.0952) > 0$$

$y = \sqrt{x}$

this is always +ve negative

this is always +ve always

$$y > 10.498$$

At $x = 108$ will below the critical value.

$$y = \sqrt{x}$$

$$x > (10.4012)^2$$

$$x > 108.19$$