Arun Hari Anand

CS 30 - HW #8

① a) $\{S, F\}^{105}$ is the sample space, where S refers to success on a given play, and F refers to failure.

b) Event $E_S = \{S\}$ and event $E_F = \{F\}$

c) $P(S) = \frac{1}{38}$, $P(F) = \frac{37}{38}$

d) Random variable X_i is a random variable whose value is 1 if the i^{th} trial is S and 0 if the i^{th} trial is a failure. As such, X_i is an indicator variable for E_S.

e) The Probability distribution X_i is: $P(X_i = 1) = \frac{1}{38}$ and $P(X_i = 0) = \frac{37}{38}$

f) Let us define a random variable $Y = \sum_{i=1}^{105} X_i$. As such Y is the number of successful trials of the 105 total trials. We know the reward we receive is $36Y$. We also know that we lose a total of 105 dollars over the course of 105 trials. So, the amount of money we have at the end is $105 - 105 + 36Y$, which must be greater than 105.

$$105 - 105 + 36Y > 105$$

$$36Y > 105 \; ; \; Y > 2.92$$

Since Y may only be a whole number, $Y \geq 3$. So we are calculating $P(Y \geq 3)$.

Since Y has a binomial since it is a sum of 105 bernoulli trials:

$$P(Y=0) = \left(\frac{37}{38}\right)^{105} ; P(Y=1) = 105\left(\frac{37}{38}\right)^{104}\left(\frac{1}{38}\right)$$

$$P(Y=2) = \binom{105}{2}\left(\frac{37}{38}\right)^{103}\left(\frac{1}{38}\right)^2$$

$$\text{So } P(Y \geq 3) = 1 - \left(\frac{37}{38}\right)^{105} - 105\left(\frac{37}{38}\right)^{104}\left(\frac{1}{38}\right) - \binom{105}{2}\left(\frac{37}{38}\right)^{103}\left(\frac{1}{38}\right)^2$$

$$= 0.5242$$

(2) Sample space: $\{R, L, M, D\} = N$. The sample space is
$S = \{s \in N^* \mid$ every element of N appears atleast once and the last element of the string appears exactly once$\}$.

$l_i(S)$ = the length of the shortest prefix of the string S such that the ith distinct element of N first appears.

$l_0(S) = 0$.

$P(R) = P(L) = P(M) = P(D) = \frac{1}{4}$.

a) Random Variable X denotes the number of times you have eaten a happy meal before you first encounter Leonardo. That is, it is the number of Happy Meals consumed until E_L occurs.

Since $P(L) = \frac{1}{4}$ and so $P(\bar{L}) = \frac{3}{4}$, this may be viewed a binomially distributed variable, with $P(X=1) = \frac{1}{4}$, $P(X=2) = \left(\frac{3}{4}\right)\left(\frac{1}{4}\right)$ and $P(X=n) = \left(\frac{3}{4}\right)^{n-1}\left(\frac{1}{4}\right)$, with x assuming all integer values such that $x \geq 1$.

$$E[X] = \sum_{x=1}^{\infty} x\left(\frac{3}{4}\right)^{n-1}\left(\frac{1}{4}\right) = \frac{1}{4}\sum_{x=1}^{\infty}(x)\left(\frac{3}{4}\right)^{x-1}$$

$$= \frac{1}{4}\left(1 + 2\left(\frac{3}{4}\right) + 3\left(\frac{3}{4}\right)^2 \dots\right) = S$$

$$1 + 2\left(\frac{3}{4}\right) + 3\left(\frac{3}{4}\right)^2 \dots = 4S$$

$$\frac{3}{4} + 2\left(\frac{3}{4}\right)^2 + 3\left(\frac{3}{4}\right)^3 = (4S)\left(\frac{3}{4}\right) = 3S$$

$$4S - 3S = 1 + (2-1)\frac{3}{4} + (3-2)\left(\frac{3}{4}\right)^2 \dots$$

$$= 1 + \frac{3}{4} + \left(\frac{3}{4}\right)^2 + \left(\frac{3}{4}\right)^3 \dots$$

$$= 1 + \sum_{i=1}^{\infty}\left(\frac{3}{4}\right)^i$$

$$= 1 + \frac{3/4}{1-(3/4)} = 1 + \frac{3/4}{1/4} = 1 + 3$$

$$= 4 = S$$

Since $S = E[X]$, the expected number of happy meals before encountering leonardo is 4.

b) Let Random Variable X_i be the number of happy meals until the first instance of the i^{th} toy is encountered, where $i=1$ if the toy is the first toy of all 4 to be in the series, $i=2$ if it is the second, and so on. As such $X_i(S) = l_i(S) - l_{i-1}(S)$

X_1: $P(X_1=1) = 1$, and $P(X_1 = x)$ for all $x \neq 1 = 0$. So $E[X_1] = 1$.
Once the first toy is encountered, the probability of encountering the second toy in a given happy meal is $\frac{3}{4}$, since there are 3 undiscovered toys. So $E[X_2] = 1/p$, where $p = \frac{3}{4}$, so $E[X_2] = 4/3$.
By the same logic, after the second toy has been found, the expected number of happy meals to be consumed before the third one is found is 2, because the probability of discovering the third toy in a given happy meal is $\frac{1}{2}$. So $E[X_3] = 2$
Also, after the third toy has been found, since the probability of finding the fourth toy in a given happy meal is $\frac{1}{4}$, $E[X_4] = 4$ by the same logic.

Let us define the random variable $Y = X_1 + X_2 + X_3 + X_4$ as the number of happy meals to be eaten before finding all 4 toys.
$E[Y] = E[X_1 + X_2 + X_3 + X_4] = E[X_1] + E[X_2] + E[X_3] + E[X_4]$, by linearity of expectation.
So $E[Y] = 1 + \frac{4}{3} + 2 + 4 = 8\frac{1}{3}$ or $\frac{25}{3}$

c) For n toys, let us assume that $Y = \sum_{i=1}^{n} X_i$, where X_i is a random variable that represents the number of happy meals that must be eaten to discover the i^{th} toy after the $(i-1)^{th}$ toy has been discovered. $X_i = l_i(S) - l_{i-1}(S)$

prob (finding the i^{th} in a given happy meal) $= \frac{n-i+1}{n}$, because we know prob(finding the 1^{st} toy) is $\frac{n}{n}$.

By the logic discussed in part b) $E[X_i] = \frac{n}{n-i+1}$

$$E[Y] = E\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} E[X_i] = \sum_{i=1}^{n} \frac{n}{n-i+1}$$

$$= \boxed{n\sum_{i=1}^{n} \frac{1}{n-i+1}, \text{ since } n \text{ is constant}}$$

3) Proof by counterexample:

$P(X=1 \wedge Y=1)$ = Prob(having one of the coins be heads and the other be tails)

Let C_1 and C_2 be the two different coins; the sample space is therefore $\{(H,H),(H,T),(T,H),(T,T)\}$, where the first number in the pair denotes whether C_1 had heads or tails, and the second number denotes whether C_2 turned up heads or tails.

Let event $E_H = \{(H,H),(H,T)\}$, $E_T = \{(T,H),(T,T)\}$, $F_H = \{(H,H),(T,H)\}$, $F_T = \{(H,T),(T,T)\}$.

E_H is the event that C_1 turns up heads, E_T is the event that C_1 turns up tails. F_H is the event that C_2 turns up heads and F_T is the event that F_1 turns up tails.

$$\begin{aligned} P(X_1=1 \wedge Y=1) &= P(E_H \cap F_T) + P(E_T \cap F_H) \\ &= P(E_H|F_T)\cdot P(F_T) + P(E_T|F_H)\cdot P(F_H) \\ &= (\tfrac{1}{2})\cdot(\tfrac{1}{2}) \quad + (\tfrac{1}{2})(\tfrac{1}{2}) \\ &= \tfrac{1}{2} \end{aligned}$$

$P(X_1=1)$ = Prob(having exactly 1 heads, and therefore 1 tails) $=\frac{1}{2}$, as calculated above. $P(X=0)=\frac{1}{4}$ and $P(X=2)=\frac{1}{4}$

$P(Y=1)$ = Prob(having exactly 1 tails and therefore 1 heads) $=\frac{1}{2}$, as calculated above. $P(Y=0)=\frac{1}{4}$ and $P(Y=2)=\frac{1}{4}$

$$P(X=1)\cdot P(Y=1) = \tfrac{1}{2}\cdot\tfrac{1}{2} = \tfrac{1}{4} \neq \tfrac{1}{2} = P(X=1 \wedge Y=1).$$

So, the events are not independent.

4) A dodecahedral pair of dice have 12 faces each, with numbers ranging from 1 to 12.

As such, the sample space is the set $[1...12] \times [1...12] = S$

X_1 is a random variable that represents the number that comes up on the first die, and X_2 is a random variable that represents the number that comes up on the second die.

X_1 and X_2 may assume all values a such that $a \in \mathbb{N}$, and $1 \le a \le 12$. $P(X_1 = a) = \frac{1}{12}$ for all a and $P(X_2 = a) = \frac{1}{12}$ for all a.

a) Let X be the random variable that represents the sum of the die rolls. As such $X = X_1 + X_2$

$$E[X] = E[X_1 + X_2] = E[X_1] + E[X_2] \text{ by linearity of expectation}$$

$$E[X_1] = \sum_{i=1}^{12} \left(\frac{1}{12}\right) i = \frac{1}{12} \sum_{i=1}^{12} i = \frac{1}{12}\left(\frac{12+1}{2}\right)(12) = \frac{12+1}{2} = \boxed{\frac{13}{2}}$$

$$E[X_2] = \sum_{i=1}^{12} \frac{1}{12} i = \frac{13}{2} \text{ by the logic above for } E[X_1]$$

$$E[X_1 + X_2] = E[X] = 13$$

b) $X = X_1 + X_2$ as in part a)

$$V[X] = V[X_1 + X_2] = V(X_1) + V(X_2) \text{ since } X_1 \text{ and } X_2 \text{ are independent}$$

$$V[X_1] = E[X_1^2] - E^2[X_1]$$

$$= \sum_{i=1}^{12} (i)^2 \left(\frac{1}{12}\right) = \frac{1}{12} \sum_{i=1}^{12} (i^2) = \frac{1}{12}\left(\frac{n(n+1)(2n+1)}{6}\right), n = 12$$

$$= \frac{1}{12}\left(\frac{12(13)(25)}{6}\right) = \frac{(13)(25)}{6} = \frac{325}{6}$$

$V(X_1) = \frac{325}{6} - \left(\frac{13}{2}\right)^2$ from part a) =

$$= \frac{325}{6} - \frac{169}{4} = \frac{143}{12}$$

$$V[X_2] = \sum_{i=1}^{12} (i)(\tfrac{1}{2}) - \left(\frac{13}{2}\right)^2, \text{ since } E[X_2^2] - E^2[X_2] = V[X_2]$$

$= \frac{143}{12}$ from the same logic as above for $V[X_1]$

$$V[X_1 + X_2] = V[X_1] + V[X_2] = \frac{143}{12} + \frac{143}{12} = \boxed{\frac{143}{6}}$$

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

Sample space is $\{S, F\}$ and $P(S) = p$ and $P(F) = q = 1-p$

E_S is the event that represents success or $\{S\}$, and E_F represents $\{F\}$

X is a random variable such that $X=1$ when event E_S occurs and $X=0$ when E_F occurs.

$P(X=0) = q = 1-p$ and $P(X=1) = p$

$E[X] = (1)(p) + 0(1-p) = p$, so $E^2[X] = p^2$.

$E[X^2] = (1)^2(p) + 0^2(1-p) = p$.

$$E[X^2] - E^2[X] = p - p^2$$

From calculus, we know that the max of $p-p^2$ can be calculated as follows:

$$\frac{d}{dp}[p-p^2] = 1-2p = 0 \text{ when } p = \tfrac{1}{2}$$

Since $\frac{d}{dp}[1-2p] = -2 < 0$, $p = \frac{1}{2}$ is a max for $p-p^2$.

So $p-p^2$ is at a max when $p = \frac{1}{2}$, and $p-p^2 = \frac{1}{2} - \frac{1}{4} = \boxed{\frac{1}{4}}$

So $p-p^2 = E[X^2] - E[X] = V[X] \leq \frac{1}{4}$.

(6) Sample space: $M = [1 \dots m]$, $S = \{ s \in M^k \mid$ every element of M appears at least once in s and the last element of s appears exactly once$\}$.

Let Y be a random variable that represents the elements needed to fill all slots of a hash table. Let $l_i(s)$ be the length of the shortest prefix of the sequence such that the i^{th} distinct slot is filled. We now define $X_i(s) = l_i(s) - l_{i-1}(s)$, which is equal to the length of the string that lies between the $(i-1)^{th}$ distinct hashing and the i^{th} distinct hashing. $l_0(s) = 0$. Y therefore is $\sum_{i=1}^{m} X_i(s)$

$$E[Y] = E\left[\sum_{i=1}^{m} X_i\right] = \sum_{i=1}^{m} E[X_i], \text{ by linearity of expectation}$$

We know that, from problem (2), since the probability of finding the i^{th} empty slot is $\frac{k-i+1}{k}$ (since the distribution is uniform),

$E[X_i] = \frac{k}{k-i+1}$, as shown in problem (2)

$$\sum_{i=1}^{k} E[X_i] = \sum_{i=1}^{k} \left(\frac{k}{k-i+1}\right) = \boxed{k \sum_{i=1}^{k} \frac{1}{k-i+1}}, \text{ since } k \text{ is a constant}$$

$$\boxed{\text{So } E[Y] = k \sum_{i=1}^{k} \frac{1}{k-i+1}}$$

⑦ Assume that $X_1 \ldots X_n$ are pairwise independent. That is, $E[X_i X_j] = E[X_i] \cdot E[X_j]$ for all $i, j \in \mathbb{N}$, $i \neq j$

As such $V\left(\sum_{i=1}^{n} X_i\right) = E\left(\left(\sum_{i=1}^{n} X_i\right)^2\right) - E^2\left(\sum_{i=1}^{n} X_i\right)$

$= E((X_1 + X_2 \ldots X_n)^2) - E^2(X_1 + X_2 \ldots X_n)$

$= E((X_1 + X_2 \ldots X_n)^2) - (E(X_1) + E(X_2) \ldots + E(X_n))^2$ by linearity of expectation

$= E\left(\sum_{i=1}^{n} \sum_{j=1}^{n} X_i X_j\right) - \sum_{i=1}^{n} \sum_{j=1}^{n} E(X_i) E(X_j)$, because $(a_1 + a_2 \ldots a_n)^2 = \sum_{i=1}^{n} \sum_{j=1}^{n} a_i a_j$

$= E\left(\sum_{i=1}^{n} X_i^2\right) + E\left(\sum_{i=1}^{n} \sum_{\substack{j=1 \\ j \neq i}}^{n} X_i X_j\right) - \left(\sum_{i=1}^{n} E^2(X_i) + \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j \neq i}}^{n} E(X_i) E(X_j)\right)$

This is done by splitting the square cases, and using linearity of expectation

$= \sum_{i=1}^{n} E(X_i^2) + \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j \neq i}}^{n} E(X_i X_j) - \sum_{i=1}^{n} E^2(X_i) - \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j \neq i}}^{n} E(X_i) E(X_j)$

using linearity of expectation.

$= \sum_{i=1}^{n} E(X_i^2) + \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j \neq i}}^{n} E(X_i) E(X_j) - \sum_{i=1}^{n} E^2(X_i) - \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j \neq i}}^{n} E(X_i) E(X_j)$

Since we know that $X_1 \ldots X_n$ are pairwise independent, and $E(X_i X_j) = E(X_i) E(X_j)$ for independent random variables X_i and X_j

$= \sum_{i=1}^{n} E(X_i^2) - \sum_{i=1}^{n} E^2(X_i)$

$= (E(X_1^2) - E^2(X_1)) + (E(X_2^2) - E^2(X_2)) + (E(X_n^2) - E^2(X_n))$

$= Var(X_1) + Var(X_2) + Var(X_n)$

$= \sum_{i=1}^{n} Var(X_i)$

Thus, we have shown that $V\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} V(X_i)$ for pairwise independent random variables $X_1 \ldots X_n$.

Sample space is arbitrary with an arbitrary probability distribution

8. We know from Chebyshev's Inequality that

$$Prob(|X - E[X]| \geq r) \leq \frac{V(X)}{r^2}$$ for an arbitrary sample space

Claim: $Prob(|X - E[X]| < r) \geq 75\%$

Assume that X is a random variable that may assume integer values, with some mean $E[X]$.

Let us consider E, the event that $|X - E[X]| \geq r$, where $r = 2\sqrt{V(X)}$

$$P(E) \leq \frac{V(X)}{4V(X)}$$ by Chebyshev's Inequality

$$P(E) \leq \frac{1}{4}$$, by algebra

$$1 - P(E) \geq 1 - \frac{1}{4}$$, since $P(E)$ is at most $\frac{1}{4}$

$$P(\bar{E}) \geq \frac{3}{4}$$, since $1 - P(E) = P(\bar{E})$

As such, $Prob(|X - E[X]| < r) \geq \frac{3}{4} = 75\%$

The claim thus follows from the direct proof principle.

(9) Let $V = [1 \dots n]$ and $E^* = \{\{u,v\} \mid u,v \in V \text{ and } u \neq v\}$.

E then is a subset of E^*, where $\{u,v\}$ denotes that there exists an edge between vertices u and v. As such, the sample space is $E \subseteq E^*$.

The probability distribution for the sample space is $p^{|E|} \cdot (1-p)^{n-|E|}$

Let us define a random variable

$$X_{\{u,v\}} = \begin{cases} 1, & \{u,v\} \in E \quad u \neq v \\ 0, & \{u,v\} \notin E \end{cases}$$

As such $P(X_{\{u,v\}} = 1) = p$ and $P(X_{\{u,v\}} = 0) = 1-p$

$E(X_{\{u,v\}}) = (p)(1) + (0)(1-p) = p.$

Let us now consider $X = \sum_{\substack{\{u,v\} \in V \\ u \neq v}} X_{\{u,v\}}$, which is the total number of

edges in the graph. As such, since $X_{\{u,v\}}$ is constant at p for all

$u \neq v$, and there are $\binom{n}{2}$ total possibilities for $\{u,v\}$,

$$E\Big(\sum_{\substack{\{u,v\} \in V \\ u \neq v}} X_{\{u,v\}}\Big) = \sum_{\substack{\{u,v\} \in V \\ u \neq v}} E(X_{\{u,v\}}) \text{ by linearity of expectation.}$$

$$= \binom{n}{2}(p)$$

b) Let $X_{\{a,b,c\}} = \begin{cases} 1, & \text{if } \{a,b\},\{b,c\},\{a,c\} \in E \text{ and } a \neq b \neq c \\ 0, & \text{if not otherwise} \end{cases}$

$P(X_{\{a,b,c\}} = 1) = \text{probability that all 3 edges exist} = p^3$

$P(X_{\{a,b,c\}} = 0) = 1 - p^3$

$E(X_{\{a,b,c\}}) = p^3 + 0(1-p^3) = p^3$

Let us now think of the sum of all such graphs isomorphic to K_3.

$$X = \sum_{\substack{\{a,b,c\} \\ a \neq b \neq c}} X_{\{a,b,c\}}$$

$$E[X] = \sum_{\substack{\{a,b,c\} \\ a \neq b \neq c}} E(X_{\{a,b,c\}}) \text{ by linearity of expectation}$$

Since there are $\binom{n}{3}$ ways to pick 3 distinct vertices from V, by the logic used in part a,

$$E[X] = \binom{n}{3} p^3$$

c) $= \binom{n}{3}\left(\frac{c}{n}\right)^3$

$= \frac{n!}{(n-3)!\,3!} \cdot \frac{c^3}{n^3} = \frac{(n)(n-1)(n-2)\,c^3}{3!\,n^3}$

$\lim_{n \to \infty} \frac{n(n-1)(n-2)c^3}{3!\,n^3}$ Since there will be a cn^3 term in the numerator and $3!n^3$ term in the denominator, and they will be the highest degree terms in the numerator and denominator respectively, the limit is $\frac{c}{3!} = \boxed{\frac{c}{6}}$