1. (a) If Monty opens a door and reveals a goat then we are left with two events in the sample space. One with chosen door has car and remaining door has computer and second with chosen door has computer and remaining door has computer. Hence chances are fifty-fifty that contestant will get car if he or she switches.

So if switched probability of getting car $= \frac{1}{2}$

(b) Let, $C_i =$ ith door has car for $i = 1,2,3$
$D_i =$ ith door has computer for $i=1,2,3$

Let contestant choses door 1 and Monty reveals door 2 has computer. So if contestant switches, we need to calculate $P(C_3|D_2)$

From Baye's theorem,

$$P(C_3|D_2) = \frac{P(D_2|C_3)\,P(C_3)}{P(D_2)}$$

$P(C_i) = \frac{1}{3}$ for $i = 1, 2, 3$

$P(D_2|C_3) = p$ (because door 2 is less preferred choice given C_3)

$P(D_2|C_2) = 0$ (because door 2 has computer from the given criteria)

$P(D_2|C_1) = q$ (Now door 2 is more preferred choice because door 3 has goat ~~since~~ given C_1)

$$\Rightarrow \frac{P(D_2|C_3)\,P(C_3)}{P(D_2)} = \frac{\frac{1}{3} \times p}{P(D_2)}$$

From LOTP,

$$\frac{\frac{1}{3} \times p}{P(D_2|C_1)\,P(C_1) + P(D_2|C_2)\,P(C_2) + P(D_2|C_3)\,P(C_3)}$$

$$= \frac{\frac{p}{3}}{\frac{p}{3} + \frac{q}{3}} = \frac{\frac{p}{3}}{\frac{p+q}{3}} = \frac{\frac{p}{3}}{\frac{1}{3}} = p$$

Hence contestant should switch if $p > \frac{1}{2}$

(2) The recurrence relation for P_i (Probability of reaching N dollars starting with i dollors) is:

$$P_i = \frac{2}{3} P_{i-1} + \frac{1}{3} P_{i+1}$$

Solution to this difference equation is

$$\frac{1-\left(\frac{q}{p}\right)^i}{1-\left(\frac{q}{p}\right)^N}, \quad \text{where } q = 2/3 \text{ and } p = \frac{1}{3}$$

Here, q/p is 2 and $N = i+2$

$$\Rightarrow P_i = \frac{1-2^i}{1-2^{i+2}} = \frac{2^i - 1}{4 \times 2^i - 1}$$

$$= \frac{1}{4}\left(\frac{2^i - 1}{2^i - \frac{1}{4}}\right)$$

Now, $2^i - \frac{1}{4} > 2^i - 1$

$$\Rightarrow \frac{2^i - 1}{2^i - \frac{1}{4}} < 1$$

$$\Rightarrow \frac{1}{4}\left(\frac{2^i - 1}{2^i - \frac{1}{4}}\right) < \frac{1}{4}$$

$$\Rightarrow P_i < \frac{1}{4}$$

3.(a)

	C_1	C_2
green	141	19
red	39	1
green %	78.33 %	95%

	M_1	M_2
green	10	158
red	10	22
green %	50%	87.7 %

overall % of green in C_1 and $C_2 = \frac{141+19}{141+19+39+1}$

$= \frac{160}{200} = 80\%$

overall % of green in M_1 and $M_2 = \frac{10+158}{10+158+10+22}$

$= \frac{168}{200} = 84\%$

(b) A is percentage of green gummi bears

B is Mauve jar and B^c is Crimson jar

C is Jar 1 and C^c is Jar 2

Hence,

$P(A|B,C) < P(A|B^c,C)$

$P(A|B,C^c) < P(A|B^c,C^c)$

but, $P(A|B) > P(A|B^c)$

Intuitively, ratio of individuals jars are compared first and then overall ratio is being compared here and results are opposite in the two cases. Hence relates to Simpson's paradox.

4. (a) We need to find $P(D \mid T_1, T_2, T_3 \ldots T_n)$

From Baye's theorem,

$$P(D \mid T_1, T_2, T_3 \ldots T_n) = \frac{P(T_1, T_2, T_3 \ldots T_n \mid D)\, P(D)}{P(T_1, T_2 \ldots T_n)}$$

Given D or D^c $T_1, T_2, T_3 \ldots T_n$ are independant.

$$\Rightarrow P(T_1, T_2, T_3 \ldots T_n \mid D) = P(T_1 \mid D)\, P(T_2 \mid D) \ldots P(T_n \mid D)$$

and

$$P(T_1, T_2, T_3 \ldots T_n \mid D^c) = P(T_1 \mid D^c)\, P(T_2 \mid D^c) \ldots P(T_n \mid D^c)$$

$$\Rightarrow P(T_1, T_2, T \ldots T_n \mid D) = a^n$$

and, $P(T_1, T_2, T_3 \ldots T_n \mid D^c) = b^n$

$$\Rightarrow P(D \mid T_1, T_2 \ldots T_n) = \frac{p a^n}{P(T_1, T_2 \ldots T_n)}$$

From LOTP,

$$P(T_1, T_2 \ldots T_n) = P(T_1, T_2 \ldots T_n \mid D)\, P(D) + P(T_1, T_2 \ldots T_n \mid D^c)\, P(D^c)$$

$$= p a^n + q b^n$$

~~$\Rightarrow P(T_1, T_2, T_3 \ldots$~~

$$\Rightarrow P(D \mid T_1, T_2 \ldots T_n) = \frac{p a^n}{p a^n + q b^n}$$

(b) From the question,

$$P(T_j \mid G) = 1 \qquad \text{for } j = 1, 2, 3 .. n$$

Now, $$P(D \mid T_1, T_2 \dots T_n) = \frac{P(T_1, T_2 \dots T_n \mid D)\, P(D)}{P(T_1, T_2 \dots T_n)} \quad \text{— (i)}$$

$$P(T_1, T_2 \dots T_n \mid D) = P(T_1, T_2 \dots T_n \mid D, G)\, P(G) + P(T_1, T_2 \dots T_n \mid D, G^c)\, P(G^c)$$

$$= 1 \times \frac{1}{2} + P(T_1 \mid D, G^c)\, P(T_2 \mid D, G^c) \dots P(T_n \mid D, G^c) \times \frac{1}{2}$$

$$= \frac{1}{2}(1 + a_0^n)$$

So numerator of (i) $= \frac{p}{2}(1 + a_0^n)$

Using LOTP,

$$P(T_1, T_2 \dots T_n) = P(T_1, T_2 \dots T_n \mid D)\, P(D) + P(T_1, T_2 \dots T_n)\, P(D^c)$$

$$P(T_1, T_2 \dots T_n \mid D)\, P(D) = \frac{p}{2}(1 + a_0^n)$$

$$P(T_1, T_2 \dots T_n \mid D^c)\, P(D^c) = q\big(P(T_1, T_2 \dots T_n \mid D^c, G)\, P(G) + P(T_1, T_2 \dots T_n \mid D^c, G^c)\, P(G^c)\big)$$

$$= \frac{q}{2}(1 + b_0^n)$$

Hence denominator of (i) $= \frac{p}{2}(1 + a_0^n) + \frac{q}{2}(1 + b_0^n)$

$$\Rightarrow P(D \mid T_1, T_2 \dots T_n) = \frac{\frac{p}{2}(1 + a_0^n)}{\frac{p}{2}(1 + a_0^n) + \frac{q}{2}(1 + b_0^n)}$$

$$= \frac{p(1 + a_0^n)}{p(1 + a_0^n) + q(1 + b_0^n)}$$

5. (a) Let M = mother has disease

C_1 = child one has disease

C_2 = child two has disease

we need to find $P(C_1^c, C_2^c)$

Using LOTP,

$$P(C_1^c, C_2^c) = P(C_1^c, C_2^c | M) P(M) + P(C_1^c, C_2^c | M^c) P(M^c)$$

$$P(M) = \frac{1}{3}, \quad P(M^c) = \frac{2}{3}$$

$$P(C_1^c, C_2^c | M^c) = 1$$

$$P(C_1^c, C_2^c | M) = P(C_1^c | M) P(C_2^c | M)$$

$$= \frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$$

$$\Rightarrow P(C_1^c, C_2^c) = \frac{1}{4} \times \frac{1}{3} + 1 \times \frac{2}{3} = \frac{1}{12} + \frac{2}{3}$$

$$= \frac{9}{12} = \frac{3}{4}$$

(b) If any child has disease implies mother has disease and thus other child may or may not have disease. Thus if one child has disease, it predicts something about disease status of second child. Hence they are not independant.

(c) $$P(M \mid c_1^c, c_2^c) = \frac{P(c_1^c, c_2^c \mid M)\, P(M)}{P(c_1^c, c_2^c)}$$

$$= \frac{\frac{1}{2} \times \frac{1}{2} \times \frac{1}{3}}{\frac{3}{4}}$$

$$= \frac{1}{12} \times \frac{4}{3} = \frac{1}{9}$$

6. (a) $P(X=k) = 1 -$ Probability of no success in k trials

$$= 1 - \left(\tfrac{1}{2}\right)^k$$

$$= \frac{2^k - 1}{2^k}$$

(b) If k trials are performed all trials will have atleast one success and atleast one failure except these two cases

(i) All trials result in failure

(ii) All trials result in success

And there 2^k possible results,

Hence, $P(X=k) = \dfrac{2^k - 2}{2^k} \quad \forall\, k \geq 2$

$$= 0 \quad \forall\, 0 < k < 2$$

(7) In the pebble world, given that x is in B we remove all the pebbles that are not in B. Now for remaining pebbles, probability of ~~any~~ every pebble is same because distribution is uniform.

Hence, $P(X = x \mid X \text{ in } B) = \frac{1}{|B|}$

$$(X \mid X \text{ in } B) \sim \text{DUnif}(B)$$

8.(a) r.v. X = number of winning tickets

$X \sim Bin(3, P)$

$P(X=k) = \binom{3}{k} P^k (1-P)^k$

~~(b)~~ ~~Probab~~

(b) Using Inclusion-exclusion,

Probability of atleast one winning ticket

$= P(X=1) + P(X=2) + P(X=3)$

$= \binom{3}{1} P(1-P)^2 + \binom{3}{2} P^2(1-P) + \binom{3}{3} P^3$

$= P^3 + 3P(1+P^2-2P) + 3P^2(1-P)$

$= P^3 + 3P + \cancel{3P^3} - 6P^2 + 3P^2 - \cancel{3P^3}$

$= P^3 + 3P - 3P^2$

$= 3P - 3P^2 + P^3$

~~(c)~~ ~~3p~~ Using Combliment,

Required probability $= 1 - P(X=0)$

$= 1 - \binom{3}{0} P^0 (1-P)^3$

$= 1-(1-P)^3 \quad = 1 - 1 + P^3 - 3P^2 + 3P$

$= 3P - 3P^2 + P^3$

(C) $3P - 3P^2 + P^3 = 3P - \{P^2(3-P)\}$

Now, $P^2(3-P) > 0$

Hence, $3P - 3P^2 + P^3 < 3P$

If P is very small then $P^2(3-P)$ tends to zero.

Inthat case required probability is close to $3P$.

9. (a) From LOTP,

$$P(X=x) = P(X=x|C_1)\,P(C_1) + P(X=x|C_2)\,P(C_2)$$

where, C_1 = coin one is chosen

C_2 = coin two is chosen

$$P(C_1) = P(C_2) = \frac{1}{2}$$

$$\Rightarrow P(X=x) = \frac{1}{2}\left(P(X=x|C_1) + P(X=x|C_2)\right)$$

Now, $P(X=x|C_1) = \binom{n}{x} P_1^x (1-P_1)^{n-x}$

and $P(X=x|C_2) = \binom{n}{x} P_2^x (1-P_2)^{n-x}$

So, $P(X=x) = \frac{1}{2}\left(\binom{n}{x} P_1^x (1-P_1)^{n-x} + \binom{n}{x} P_2^x (1-P_2)^{n-x}\right)$

(b) Let $P_1 = P_2 = P$

$$\Rightarrow P(X=x) = \frac{1}{2}\left(\binom{n}{x} P^x (1-P)^{n-x} + \binom{n}{x} P^x (1-P)^{n-x}\right)$$

$$= \binom{n}{x} P^x (1-P)^{n-x}$$

Hence, $X \sim \text{Bin}(n, P)$

where $P = P_1 = P_2$

(c) If $P_1 \neq P_2$ then each coin toss is not a indepedant Bernouli trial because previous trials will predict outcomes of future coin tosses as it will give us hint about which coin is chosen. Hence each trial is not an independant Bernoulli trials which is a requirement of Binomial distribution.

(10) Any given person will vote for kodos with probability $P = P_1 P_2 P_3$

Each person will either vote for kodos or will not vote for kodos and each person's vote is independant of each other.

r.v. X: number of votes for kodos

X has Binomial distribution

$X \sim Bin(n, P)$

$P(X = x) = \binom{n}{x} P^{x} (1-P)^{n-x}$

where $P = P_1 P_2 P_3$

11.(a) If first two tosses are heads out of $X = x$ ~~t~~heads then remaining $(n-2)$ tosses will result in $x-2$ heads and distribution will be $Bin(n-2, \frac{1}{2})$

Let F = first two heads

$$P(X = x \mid F) = \binom{n-2}{x-2} \left(\frac{p}{2}\right)^{x-2} \left(\frac{1}{2}\right)^{n-\cancel{x}2-x-2}$$

$$= \binom{n-2}{x-2} \left(\frac{1}{2}\right)^{n-2} = \binom{8}{x-2} \left(\frac{1}{2}\right)^{8}$$

(b) $P(X = x) = 0$ for $0 \leq x < 2$

otherwise, $P(X = x) = \binom{n}{x} \left(\frac{1}{2}\right)^{n}$ (because if number of heads are atleast two then for $x \geq 2$ its ~~true~~ anyways)

12. (a) From the question

$$P(X=x) = \binom{n}{x} p^{x}(1-p)^{n-x}$$

Let Y is a r.v.

Y = Number of statistics major in sample size m

For a given $X=x$ if we have to chose $Y=y$ statistics major then its a hypergeometric distribution.

~~$\Rightarrow P(Y=y \mid X=x) =$~~

$$(Y \mid X=x) \sim HGeom(x, n-x, m)$$

And from Law of total probability,

$$P(Y=y) = \sum_{x=y}^{n} P(Y=y \mid X=x)\, P(X=x)$$

$$= \sum_{x=y}^{n} \frac{\binom{x}{y}\binom{n-x}{m-y}}{\binom{n}{m}} \binom{n}{x} p^{x}(1-p)^{n-x}$$

(b) Once we have sample drawn of sizen m, then amongst those m students each student has same probability of being a statistics major student. And they are independantly either a statistics students or not. Hence a Binomial distribution.

$$Y \sim Bin(m, p)$$

$$\Rightarrow P(Y=y) = \binom{m}{y} p^{y}(1-p)^{y}$$