Homework 2

1) x is a vector

> length (x)

Returns the length of the vector x, i.e. counts the number of elements in x.

> x [2]

If the length of $x \geq 2$ then this returns the index 2 element of x otherwise it returns nothing.

> x [-2]

If the length of $x = 1$ then returns the value of x.
If length of $x \geq 2$ then returns the index (-2) i.e. the second last element of the vector x.

> x [1:5]

This returns the first 5 elements of x but if the length of x is less than 5 then for the remaining positions it returns NA.

> x [(length(x) -5): length(x)]

If length(x) >= 5 then returns the elements from index length(x) - 5 till the last element otherwise error.

> x [c (1, 3, 5)]

Returns the 1st, 3rd and 5th element of x. If any element is not there in x then returns NA.

> x [x > 3]

Returns the elements of x that have value greater than 3.
If none element is greater than 3 then returns numeric(0).

> x [x < -2 | x > 2]

Returns those elements of x which have a value greater than or equal to 2 otherwise numeric (0).

```
> which ( x == max(x) )
```

Returns the index of the element with maximum value.

2.a) The dataset has 10 attributes for the carat, cut, colour, clarity, depth, table, price and the size of the diamond. It has details of about 53,940 diamonds.

2.b) The categories are = carat, cut, colour, clarity, depth, table, price and the coordinates x, y, z defining the dimensions (length, width & depth) of a diamond.

2.c) The (i) code plots the variable 'count' against the variable 'cut' with a third variable 'clarity' behaving as an aesthetic.

Code (ii) also does a similar thing as in (i), the only difference being that now 'clarity' is taken vertically. In (i) 'clarity' was plotted horizontally with different colours but in (ii) it is plotted vertically.

3.a) The dataset consists of 29 observations and 3 variables density, density 2 and density 3. 'density' contains the value for the mean density of the earth. 'density 2' and 'density 3' are same as 'density' with ~~others~~ some alterations.

3.b) summary (Cavendish $ density)

output :

Min	1st Qu.	Median	Mean	3rd Qu.	Max
4.880	5.300	5.460	5.448	5.610	5.850

summary (Cavendish $ density 2)

Output :	Min	1st Qu.	Median	Mean	3rd Qu.	Max
	5.070	5.340	5.470	5.482	5.620	5.880

summary (Cavendish $ density 3)

Output :	Min	1st Qu.	Median	Mean	3rd Qu.	Max	NA's
	5.100	5.340	5.460	5.483	5.625	5.850	6

4) We roll a dice five times,

Range (Y) = {5, 6, 7, ... 30}

Let X_i be the output of roll i

$\therefore f_Y(y) = P(Y=y) = P(X_1 + X_2 + X_3 + X_4 + X_5 = y)$.

where, $1 \le X_i \le 6 \quad \forall \quad 1 \le i \le 5$

When $\underline{y=5}$,

$f_Y(5) = P(Y=5) = P(X_1 + \dots X_5 = 5)$

The only solution of the equation $X_1 + \dots X_5 = 5$ is

$X_i = 1 \quad \forall \quad 1 \le i \le 5$

$\therefore$ no. of solutions = 1

Now when $\underline{y=6}$ we need to find the number of solutions of $X_1 + \dots X_5 = 6$, $1 \le X_i \le 6 \quad \forall i$

Substitute, $X_i = z_i + 1 \quad \forall \quad 1 \le i \le 5$ to get

$z_1 + 1 + z_2 + 1 \dots + z_5 + 1 = 6, \quad 0 \le z_i \le 5$

$\Rightarrow z_1 + \dots z_5 = 1$

The no. of solutions of this eq $= \binom{1+5-1}{5-1} = {}^5C_4 = 5$

$\Rightarrow$ No. of solutions of $X_1 + \dots X_5 = 6$ is also 5.

In general, we need to find the number of solutions of,

$X_1 + X_2 + \dots X_5 = y, \quad 1 \le X_i \le 6, \quad 5 \le y \le 30$.

Let $X_i = z_i + 1 \quad \forall \quad 1 \le i \le 5$ so the eq becomes,

$z_1 + 1 + \dots z_5 + 1 = y \quad 0 \le z_i \le 5$

$\Rightarrow z_1 + \dots z_5 = y - 5$.

When $\underline{y=7}$, $z_1 + z_2 + \dots z_5 = 2$

no. of solutions $= \binom{2+5-1}{5-1} = {}^6C_4 = 15$

When $\underline{y=8}$, $z_1 + z_2 + \dots z_5 = 3$

no. of solutions $= \binom{3+5-1}{5-1} = {}^7C_4 = 35$

When $\underline{y = 9}$, $z_1 + \cdots z_5 = 4$

no. of solutions $= \binom{4+5-1}{5-1} = {}^{8}C_4 = 70$

When $\underline{y = 10}$, $z_1 + \cdots z_5 = 5$

no. of solutions $= \binom{5+5-1}{5-1} = {}^{9}C_4 = 126$

When $\underline{y = 11}$, $z_1 + \cdots z_5 = 6$

no. of solutions $= \binom{6+5-1}{5-1} = {}^{10}C_4 = 210$

But in this case we might have $z_1 = 6$ & $z_i = 0$ $\forall\, 2 \le i \le 5$

we have the constraint, $0 \le z_i \le 5$ $\forall i$

so we have to exclude the cases when $z_i \geq 6$.

Let A_i be the event when $z_i \geq 6$.

$\therefore A_1 \cup A_2 \cdots \cup A_5$ is the event when either of $z_1, z_2 \cdots$ or z_5 is greater than or equal to 6.

By the principle of inclusion and exclusion,

$$|A_1 \cup \cdots \cup A_5| = \sum_{i=1}^{5} |A_i| + (-1) \sum_{1 \le i, j \le 5} |A_i \cap A_j| + \cdots$$

$$\cdots (-1)^4 |A_1 \cap \cdots A_5|$$

for, $z_1 + \cdots z_5 = 6$, we cant have the cardinality of $A_i \cap A_j$ greater than 0 since that would imply $z_i \geq 6$ & $z_j \geq 6 \Rightarrow z_i + z_j \geq 12$

$\Rightarrow z_1 + \cdots z_5 \geq 12$ not possible.

$$\therefore |A_1 \cup \cdots \cup A_5| = \sum_{i=1}^{5} |A_i|$$

If $z_1 \geq 6$, use $w_1 = z_1 - 6$ so we have,

$w_1 + 6 + z_2 + z_3 + z_4 + z_5 = 6$

$\Rightarrow w_1 + z_2 + \cdots z_5 = 0$.

the only solution of this is, $w_1 = z_2 = \cdots z_5 = 0$.

$\Rightarrow z_2 = z_3 = \cdots z_5 = 0$ & $z_1 = 6$.

$\Rightarrow |A_1| = 1$

Similarly, $|A_i| = 1 \quad \forall i$

$\therefore |A_1 \cup \cdots \cup A_5| = 5 \times 1 = 5$

No. of solutions of $z_1 + \cdots z_5 = 6$, $0 \leq z_i \leq 5$

is $210 - 5 = 205$.

When, $\underline{y = 12}$, $z_1 + \cdots z_5 = 7$

In this case also we might have some $z_i \geq 6$ so

proceeding as before we get,

$w_1 = z_1 - 6$,

$w_1 + 6 + z_2 + \cdots z_5 = 7$

$\Rightarrow w_1 + z_2 + \cdots z_5 = 1$

no of solutions of this equation $= \binom{1+5-1}{5-1} = 5$

$\Rightarrow |A_i| = 5 \quad \forall i$

Also, $A_i \cap A_j = \phi$ as before.

$\therefore |A_1 \cup \cdots \cup A_5| = 5 \times |A_i| = 25$

The no. of solutions of $z_1 + \cdots z_5 = 7$ is

$= \binom{7+5-1}{5-1} - 25 = 330 - 25 = 305$

we continue as above.

Now when $\underline{y=13}$, the no. of solutions of $z_1 + \cdots z_5 = 8$

$$= \binom{8+5-1}{5-1} - 5\binom{8-6+5-1}{5-1}$$

$$= 495 - 5 \times 15 = 420$$

Now when $\underline{y=14}$, the no. of solutions of $z_1 + \cdots z_5 = 9$

$$= \binom{9+5-1}{5-1} - 5\binom{9-6+5-1}{5-1}$$

$$= 715 - 5 \times 35$$

$$= 540 .$$

when $\underline{y=15}$, the no. of solutions of $z_1 + \cdots z_5 - 10$

$$= \binom{10+5-1}{5-1} - 5\binom{10-6+5-1}{5-1}$$

$$= 1001 - 5 \times 70 = 651$$

when $\underline{y=16}$, the no. of solutions of $z_1 + \cdots z_5 = 11$

$$= \binom{11+5-1}{5-1} - 5\binom{11-6+5-1}{5-1}$$

$$= 1365 - 5 \times 126 = 735$$

when $\underline{y=17}$, we need to find the no. of solutions of

$$z_1 + \cdots z_5 = 12, \quad 0 \le z_i \le 5 .$$

In this case we might have $z_i \geq 6$ & $z_j \geq 6$

$\Rightarrow A_i \cap A_j \neq \phi$.

But we can't have $z_i \geq 6$ & $z_j \geq 6$ & $z_k \geq 6$ because

$\Rightarrow z_i + z_j + z_k \geq 18$

$\Rightarrow z_1 + \cdots z_5 \geq 18$ not possible .

Let us compute $|A_1 \cap A_2|$ i.e. when $x_1 \geq 6$ & $x_2 \geq 6$.

Let $w_1 = x_1 - 6$, $w_2 = x_2 - 6$. Our equation becomes,

$w_1 + 6 + w_2 + 6 + x_3 + x_4 + x_5 = 12$

$\Rightarrow w_1 + w_2 + x_3 + x_4 + x_5 = 0$. $0 \leq w_1, w_2, x_3, x_4, x_5$

The only possible solution of this eq^n is,

$w_1 = w_2 = x_3 = x_4 = x_5 = 0$.

$\Rightarrow |A_1 \cap A_2| = 1 = |A_i \cap A_j| \quad \forall\, 1 \leq i, j \leq 5$.

$\therefore |A_1 \cup \ldots \cup A_5| = 5 \cdot |A_1| - {}^5C_2\, |A_1 \cap A_2|$

$= 5\binom{12-6+5-1}{5-1} - {}^5C_2\binom{12-6\cdot 2+5-1}{5-1}$

$= 5 \times 210 - 10 \times 1 = 1040$.

No. of solutions of $x_1 + \ldots x_5 = 12 = \binom{12+5-1}{4} - 1040$

$= 1820 - 1040 = 780$

When $\underline{y = 18}$, no. of solutions of $x_1 + \ldots x_5 = 13$

$= \binom{13+5-1}{5-1} - 5\binom{13-6+5-1}{5-1} + {}^5C_2\binom{13-2\cdot 6+5-1}{5-1}$

$= 2380 - 5 \times 330 + 10 \times 5 = 780$

When $\underline{y = 19}$, no. of solutions of $x_1 + \ldots x_5 = 14$

$= \binom{14+5-1}{5-1} - 5\binom{14-6+5-1}{5-1} + {}^5C_2\binom{14-2\cdot 6+5-1}{5-1}$

$= 3060 - 5 \times 495 + 10 \times 15 = 735$

When $\underline{y=20}$, no. of solutions of $z_1 + \cdots z_5 = 15$

$$= \binom{15+5-1}{5-1} - 5\binom{15-6+5-1}{5-1} + {}^5C_2\binom{15-2.6+5-1}{5-1}$$

$$= 3876 - 5 \times 715 + 10 \times 35 = 651$$

When $\underline{y=21}$, no. of solutions of $z_1 + \cdots z_5 = 16$

$$= \binom{16+5-1}{5-1} - 5\binom{16-6+5-1}{5-1} + {}^5C_2\binom{16-2.6+5-1}{5-1}$$

$$= 4845 - 5 \times 1001 + 10 \times 70 = 540$$

When $\underline{y=22}$, no. of solutions of $z_1 + \cdots z_5 = 17$

$$= \binom{17+5-1}{5-1} - 5\binom{17-6+5-1}{5-1} + {}^5C_2\binom{17-2.6+5-1}{5-1}$$

$$= 5985 - 5 \times 1365 + 10 \times 126 = 420$$

When $\underline{y=23}$, we need to find the no. of solutions of $z_1 + z_2 + \cdots z_5 = 18$ but in this case, $(A_i \cap A_j \cap A_k) \neq \phi$ always. Recursively we have,

$$|A_1 \cap A_2 \cap A_3| = \binom{18-3.6+5-1}{5-1}$$

$\therefore$ No. of solutions of $z_1 + \cdots z_5 = 18$

$$= \binom{18+5-1}{5-1} - 5\binom{18-6+5+1}{5-1} + {}^5C_2\binom{18-2.6+5-1}{5-1}$$

$$- {}^5C_3\binom{18-3.6+5-1}{5-1}$$

$$= 7315 - 5 \times 1820 + 10 \times 210 - 10 \times 1$$

$$= 305$$

When $\underline{y=24}$, no. of solutions of $z_1 + \cdots z_5 = 19$

$$= \binom{19+5-1}{5-1} - 5\binom{19-6+5-1}{5-1} + {}^5C_2\binom{19-2.6+5-1}{5-1} - {}^5C_3\binom{19-3.6+5-1}{5-1}$$

$$= 8855 - 5 \times 2380 + 10 \times 330 - 10 \times 5$$

$$= 205$$

When $\underline{y=25}$, no. of solutions of $z_1 + \cdots z_5 = 20$

$$= \binom{20+5-1}{5-1} - 5\binom{20-6+5-1}{5-1} + {}^5C_2\binom{20-2.6+5-1}{5-1} - {}^5C_3\binom{20-3.6+5-1}{5-1}$$

$$= 10626 - 5 \times 3060 + 10 \times 495 - 10 \times 15$$

$$= 126$$

When $\underline{y=26}$, no. of solutions of $z_1 + \cdots z_5 = 21$

$$= \binom{21+5-1}{5-1} - 5\binom{21-6+5-1}{5-1} + {}^5C_2\binom{21-2.6+5-1}{5-1} - {}^5C_3\binom{21-3.6+5-1}{5-1}$$

$$= 12650 - 5 \times 3876 + 10 \times 715 - 10 \times 35$$

$$= 70$$

When $\underline{y=27}$, no. of solutions of $z_1 + \cdots z_5 = 22$

$$= \binom{22+5-1}{5-1} - 5\binom{22-6+5-1}{5-1} + {}^5C_2\binom{22-2.6+5-1}{5-1} - {}^5C_2\binom{22-3.6+5-1}{5-1}$$

$$= 14950 - 5 \times 4845 + 10 \times 1001 - 10 \times 70$$

$$= 35$$

When $\underline{y=28}$, no. of solutions of $z_1 + \cdots z_5 = 23$

$$\binom{23+5-1}{5-1} - 5\binom{23-6+5-1}{5-1} + {}^5C_2\binom{23-2.6+5-1}{5-1} - {}^5C_3\binom{23-3.6+5-1}{5-1}$$

$$= 17550 - 5 \times 5985 + 10 \times 1365 - 10 \times 126$$

$$= 15$$

When $\underline{y=29}$, we need to find the solutions to

$$z_1 + z_2 + \cdots z_5 = 24$$

In this case it might happen, $z_i, z_j, z_k, z_l \geq 6$.

$$\Rightarrow |A_i \cap A_j \cap A_k \cap A_l| \neq 0.$$

$\therefore$ The no. of solutions =

$$\binom{24+5-1}{5-1} - {}^5C_1\binom{24-6+5-1}{5-1} + {}^5C_2\binom{24-2\cdot6+5-1}{5-1}$$

$$- {}^5C_3\binom{24-3\cdot6+5-1}{5-1} + {}^5C_4\binom{24-4\cdot6+5-1}{5-1}$$

$$= 20475 - 5\times7315 + 10\times1820 - 10\times210 + 5\times1$$

$$= 5$$

When $\underline{y=30}$, the only solution is $z_i = 6 \ \forall\ 1 \leq i \leq 5$.

$\therefore$ We shall draw a table for the no. of solutions of the equation $z_1 + \cdots z_5 = y-5$ for various values of y.

y	5	6	7	8	9	10	11	12	13	14	15	16	17
#	1	5	15	35	70	126	205	305	420	540	651	735	780

y	18	19	20	21	22	23	24	25	26	27	28	29	30
#	780	735	651	540	420	305	205	126	70	35	15	5	1

$$f_Y(y) = P(Y=y) = \frac{\#\text{ of solutions of } x_1 + \cdots x_5 = y}{\sum_{i=5}^{30} \#\text{ of solutions of } x_1 + \cdots x_5 = i}$$

5) $X \rightarrow$ outcome of the toss of a coin

$Y \rightarrow$ outcome of the roll of the die.

(a) Range (X) : $\{H, T\}$ where $H \rightarrow$ Head, $T \rightarrow$ Tail

Range (Y) : $\{1, 2, 3, 4, 5, 6\}$.

$$P(Y=1 \mid X=\text{Head}) = \frac{1}{4} = P(Y=6 \mid X=\text{Head})$$

$$P(Y=2 \mid X=\text{Head}) = P(Y=3 \mid X=\text{Head}) = P(Y=4 \mid X=\text{Head})$$

$$= P(Y=5 \mid X=\text{Head}) = \frac{1}{8}.$$

$$\therefore P(Y=y \mid X=H) = \begin{cases} \frac{1}{4} & \text{if } y=1, 6 \\ \frac{1}{8} & \text{if } y=2,3,4,5 \end{cases}$$

(b) $P(Y=1 \mid X=T) = P(Y=2 \mid X=T)$

$$= P(Y=5 \mid X=T) = P(Y=6 \mid X=T) = \frac{1}{8}$$

$$P(Y=3 \mid X=\text{Tail}) = P(Y=4 \mid X=\text{Tail}) = \frac{1}{4}$$

$$\therefore P(Y=y \mid X=\text{Tail}) = \begin{cases} \frac{1}{4} & \text{if } y = 3, 4 \\ 1/8 & \text{if } y=1,2,5,6. \end{cases}$$

(c) $$P(X=\text{Head} \mid Y=3) = \frac{P(Y=3 \mid X=\text{Head}) P(X=\text{Head})}{P(Y=3 \mid X=\text{Head}) . P(X=\text{Head}) + P(Y=3 \mid X=\text{Tail}) . P(X=\text{Tail})}$$

We get the above equation by Bayes Theorem.

Note that, $P(X=\text{Head}) = P(X=\text{Tail}) = 1/2$.

$$\therefore P(X=\text{Head} \mid Y=3) = \frac{1/8 \times 1/2}{1/8 \times 1/2 + 1/4 \times 1/2} = \frac{1}{3}$$

Book-Keeping Exercises

Ex 1.1.3 (a) The first marble drawn is white

$\Rightarrow$ The first marble is neither red nor green

$\Rightarrow$ The event is $R_1^c \cap G_1^c$

(b) The first marble drawn is green & the 2nd marble is not white

$\Rightarrow$ The first marble is green & 2nd is either red or green

$\Rightarrow$ $G_1 \cap (R_2 \cup G_2)$ is the event.

(c) If E & F are disjoint then $E \cap F = \phi$.

$E \cap F = (R_1 \cup G_2) \cap (R_1^c \cap R_2)$

$= (R_1 \cap (R_1^c \cap R_2)) \cup (G_2 \cap (R_1^c \cap R_2))$ (Distributivity)

$= (R_1 \cap R_1^c \cap R_2) \cup (R_1^c \cap G_2 \cap R_2)$ (Associativity, Commutativity)

$= (\phi \cap R_2) \cup (R_1^c \cap G_2 \cap R_2)$

$= \phi \cup (R_1^c \cap G_2 \cap R_2) = (R_1^c \cap G_2 \cap R_2)$

If $x \in R_1^c \cap G_2 \cap R_2 \Rightarrow$ The first ball is not red and the second ball is both green & red.

But the 2nd ball can either be one of red, green or white, it can't be both red & green. $\Rightarrow R_1^c \cap G_2 \cap R_2 = \phi$

$\Rightarrow E \cap F = \phi$

$\Rightarrow$ E & F are disjoint.

Ex 1.2.12 (a) There are two people. Person 1 has 365 many choices for his/her birthday. Person 2 has to have the same birthday as Person 1 so he has only 1 choice.

$\therefore |E| = 365 \times 1 = 365$

Let S_2 denote the sample space of birthdays of two people. $|S_2| = 365 \times 365$

$$\therefore P(E) = \frac{365}{365 \times 365} = \frac{1}{365} \approx 0.0027$$

(b) $F \to$ event that at most two of the three people have the same birthady.

F^c is the event that none of the three people have the same birthday with each other.

Person 1 has 365 many choices, Person 2 has 364 many choices after Person 1's birthday has been chosen &

Person 3 has 363 many choices i.e. the choices remaining after Person 1 & Person 2 have chosen their birthdays.

$\therefore |F^c| = 365 \times 364 \times 363 \Rightarrow |F| = (365)^3 - (365 \times 364 \times 363)$

$$P(F^c) = \frac{365 \times 364 \times 363}{(365)^3}$$

$$P(F) = 1 - \frac{365 \times 364 \times 363}{(365)^3} \approx 0.0082$$

(c) For a group of 4 people, let G be the event that at least two of the four have the same birthday.

Like above, $|G^c| = \underline{365 \times 364 \times 363 \times 362}$

$$\therefore P(G) = 1 - P(G^c) = 1 - \frac{365 \times 364 \times 363 \times 362}{(365)^4}$$

$$\approx 0.016$$

(d) For a group of n people, Person i has $365-(i-1)$ many choices for his birthday so that his birthday is not the same as Person 1, Person 2, .. Person $(i-1)$

If A is the event that at least two of the n people have the same birthday then,

$A^c \rightarrow$ none of the n people have the same birthday.

$\therefore |A^c| = 365 \times 364 \times \cdots 365-(n-1)$

$$P(A^c) = \frac{365 \times \cdots (365-n+1)}{(365)^n}$$

If $P(A) > 1/2 \Rightarrow$ It is more likely that at least two of the n people have the same birthday.

We can check for $n = 23$, $P(A^c) = \frac{365 \times \cdots 343}{(365)^{23}}$

$$\approx 0.493$$

$P(A) \approx 0.507$

$\therefore$ Size of the group $= 23$.

Ex 1.3.9 $X \rightarrow$ event that a die is rolled

$Y \rightarrow$ ~~even~~ number of heads.

We need to find, $P(X=5 \mid Y=5)$.

$$= \frac{P(Y=5 \mid X=5) \cdot P(X=5)}{P(Y=5)}.$$

$P(Y=5) = P(Y=5 \mid X=5)\, P(X=5) + P(Y=5 \mid X=6) P(X=6)$

$$= {}^5C_5\left(\frac{1}{2}\right)^5\left(\frac{1}{2}\right)^0 \times \frac{1}{6} + {}^6C_5\left(\frac{1}{2}\right)^5\left(\frac{1}{2}\right)^1 \times \frac{1}{6}$$

$$= \left(\frac{1}{2}\right)^5 \frac{1}{6} + \cancel{6} \times \left(\frac{1}{2}\right)^6 \times \frac{1}{\cancel{6}}$$

$$= \left(\frac{1}{2}\right)^5\left(\frac{1}{6}+\frac{1}{2}\right) = \left(\frac{1}{2}\right)^5 \times \frac{2}{3}$$

$$\therefore P(X=5 \mid Y=5) = \frac{{}^5C_5\,\cancel{(1/2)^5}\,(1/2)^0 \times \frac{1}{6}}{\cancel{(1/2)^5} \times 2/3}$$

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

Ex 1.3.10 : Let A be the event that Hannah A organises a study group & B be the event that Hannah B organises a study group.

Let O → study group is at the coffee shop

C → study group is at the campus.

$P(A) = 0.75$, $P(B) = 0.25$

$P(C \mid A) = 0.80$, $P(O \mid A) = 0.20$

$P(C \mid B) = 0.10$, $P(O \mid B) = 0.90$.

$$P(O) = P(O \mid A)P(A) + P(O \mid B)P(B)$$

$$= 0.20 \times 0.75 + 0.90 \times 0.25$$

$$= 0.375$$

$$P(A \mid O) = \frac{P(O \mid A)P(A)}{P(O)} = \frac{0.20 \times 0.75}{0.375} = 0.4$$

$$P(B \mid O) = \frac{P(O \mid B)P(B)}{P(O)} = \frac{0.90 \times 0.25}{0.375} = 0.6$$

The note says that the study group is in the coffee shop. So $A|O$ is the event that Hannah A has written the note & $B|O$ is the event that Hannah B has written the note given the study group is in the coffee shop.

$P(B|O) > P(A|O)$

$\Rightarrow$ It is more likely that Hannah B has written the note.

Ex 1.3.13: Let R_j be the event of drawing a red ball at the j^{th} step & B_j of drawing a black ball at the j^{th} step.

Total balls $= b + r$

$$P(R_1) = \frac{r}{b+r}, \quad P(B_1) = \frac{b}{b+r}$$

(a) $P(R_2) = P(R_2|R_1)P(R_1) + P(R_2|B_1)P(B_1)$

$$= \frac{r+c}{b+r+c+d} \cdot \frac{r}{b+r} + \frac{r+d}{b+r+c+d} \cdot \frac{b}{b+r}$$

$$= \frac{r(r+c) + b(r+d)}{(b+r)(b+r+c+d)}$$

(b) $P(B_2) = P(B_2|R_1)P(R_1) + P(B_2|B_1)P(B_1)$

$$= \frac{b+c}{b+r+2c} \cdot \frac{r}{b+r} + \frac{b+c}{b+r+2c} \cdot \frac{b}{b+r}$$

$$= \frac{b+c}{b+r+2c}\left(\frac{r}{b+r} + \frac{b}{b+r}\right)$$

$$= \frac{b+c}{b+r+2c}$$

(c) $P(B_3) = P(B_3|B_2)P(B_2) + P(B_3|R_2)P(R_2)$

$B_3|B_2$ is the event that the second ball drawn was black & the third is also black. At the 2nd stage we have $b+c$ black balls & $r+c$ red balls. After the second stage we get, $b+c+c$ black balls & $r+c+c$ red balls.

$$\therefore P(B_3|B_2) = \frac{b+2c}{b+2c+r+2c} = \frac{b+2c}{b+r+4c}.$$

$$P(B_3) = \frac{b+2c}{b+r+4c} \cdot P(B_2) + \frac{b+2c}{b+r+4c}(1-P(B_2))$$

$$= \frac{b+2c}{b+r+4c}.$$

Claim: At n^{th} stage there $b+(n-1)c$ many black balls & $r+(n-1)c$ many red balls.

We shall proceed by induction on n,

for $n=1$ we had b black balls & r red balls.

$n=2$, $b+c$ black balls & $r+c$ red balls.

Let it hold for $n=k-1$, i.e. at $(k-1)^{th}$ stage there $b+(k-1-1)c$ black balls & $r+(k-1-1)c$ red balls.

At the k^{th} stage, whichever ball is taken out, we add c many balls to each colour.

$\Rightarrow$ $b+(k-1-1)c+c$ black balls & $r+(k-1-1)c+c$ red balls

$\Rightarrow$ $b+(k-1)c$ black balls & $r+(k-1)c$ red balls at k^{th} step.

$$P(B_n) = P(B_n \mid B_{n-1})P(B_{n-1}) + P(B_n \mid R_{n-1})P(R_{n-1})$$

$$= \frac{b+(n-1)c}{b+r+2(n-1)c} \cdot P(B_{n-1}) + \frac{b+(n-1)c}{b+r+2(n-1)c} \cdot (1-P(B_{n-1}))$$

$$= \frac{b+(n-1)c}{b+r+2(n-1)c}.$$

(d) $c>0$, $d=0$.

$$P(B_2) = P(B_2 \mid B_1)P(B_1) + P(B_2 \mid R_1)P(R_1)$$

$B_2 \mid B_1$ is the event that given that the 1st ball was black, 2nd ball is also black.

If 1st ball was black then now we have $b+c$ many black balls & r many red balls.

Whereas for $B_2 \mid R_1$, if 1st ball is red then we have $r+c$ red balls & b black balls.

$$\therefore P(B_2) = \frac{b+c}{b+r+c} \cdot \frac{b}{b+r} + \frac{b}{b+r+c} \cdot \frac{r}{b+r}$$

$$= \frac{b}{\cancel{(b+r+c)}(b+r)} \left(\cancel{b+c+r} \right) = \frac{b}{b+r}$$

$$P(B_3) = P(B_3 | R_2)\, P(R_2) + P(B_3 | B_2)\, P(B_2)$$

$$= P(B_3 | R_2 \cap R_1)\, P(R_2 | R_1) \cdot P(R_2) + P(B_3 | R_2 \cap B_1)\, P(R_2 | B_1)\, P(B)$$

$$+ P(B_3 | B_2 \cap B_1)\, P(B_2 | B_1)\, P(B_1) + P(B_3 | B_2 \cap R_1)\, P(B_2 | R_1)\, P(R_1)$$

$$= \frac{b}{b+r+2c} \cdot \frac{r+c}{b+r+c} \cdot \frac{r}{r+b} + \frac{b+c}{b+r+2c} \cdot \frac{r}{b+r+c} \cdot \frac{b}{b+r}$$

$$+ \frac{b+2c}{b+r+2c} \cdot \frac{b+c}{b+r+c} \cdot \frac{b}{b+r} + \frac{b+c}{b+r+2c} \cdot \frac{b}{b+r+c} \cdot \frac{r}{b+r}$$

$$= \frac{br}{\cancel{(b+r+2c)}} \cdot \frac{\cancel{(b+r+2c)}}{(b+r+c)(r+b)} + \frac{b(b+c)}{\cancel{(b+r+2c)}} \cdot \frac{\cancel{b+2c+r}}{(b+r+c)(b+r)}$$

$$= \frac{b\cancel{(r+b+c)}}{\cancel{(b+r+c)}(b+r)} = \frac{b}{b+r}.$$

claim: $P(B_n) = \dfrac{b}{b+r}$

We shall proceed by induction on n.

We have seen this hold for $n = 1, 2, 3$

Let t_{n-1} be the total no. of balls in stage $n-1$.

Let our claim hold for $n-1$ i.e. $P(B_{n-1}) = \dfrac{b}{b+r}$

$$P(B_{n-1}) = \frac{\text{no. of black balls in stage } (n-1)}{\text{total no. of ball at stage } (n-1)}$$

$$\Rightarrow P(B_{n-1})\, t_{n-1} = \text{no. of black balls in stage } (n-1)$$

$$P(R_{n-1}) = \frac{\text{no. of red balls in stage } (n-1)}{t_{n-1}}$$

$$\Rightarrow P(R_{n-1})\, t_{n-1} = \text{no. of red balls in stage } (n-1)$$

Note that if we draw a red ball in stage $(n-1)$ then we have $P(R_{n-1}) \cdot t_{n-1} + c$ many red balls in stage n.

If we draw a black ball in stage $(n-1)$ then we have $P(B_{n-1}) \cdot t_{n-1} + c$ many black balls in stage n.

$$\therefore P(B_n) = P(B_n | B_{n-1}) P(B_{n-1}) + P(B_n | R_{n-1}) P(R_{n-1})$$

$$= \frac{P(B_{n-1}) t_{n-1} + c}{t_{n-1} + c} \cdot \frac{b}{b+r} + \frac{P(B_{n-1}) t_{n-1}}{t_{n-1} + c} \left(1 - P(B_{n-1})\right)$$

$$= \frac{P(B_{n-1}) t_{n-1} + c}{t_{n-1} + c} \cdot \frac{b}{b+r} + \frac{P(B_{n-1}) t_{n-1}}{t_{n-1} + c} - \frac{b}{b+r} \cdot \frac{P(B_{n-1}) t_{n-1}}{t_{n-1} + c}$$

$$= \frac{P(B_{n-1}) t_{n-1}}{t_{n-1} + c} + \frac{b}{b+r} \left[\frac{P(B_{n-1}) t_{n-1} + c}{t_{n-1} + c} - \frac{P(B_{n-1}) t_{n-1}}{t_{n-1} + c} \right]$$

$$= \frac{P(B_{n-1}) t_{n-1}}{t_{n-1} + c} + \frac{b}{b+r} \left[\frac{c}{t_{n-1} + c} \right]$$

$$= \frac{b\, t_{n-1}}{(b+r)(t_{n-1} + c)} + \frac{b c}{(b+r)(t_{n-1} + c)}$$

$$= \frac{b \cancel{(t_{n-1} + c)}}{(b+r) \cancel{(t_{n-1} + c)}} = \frac{b}{b+r}.$$

$$\Rightarrow P(B_n) = \frac{b}{b+r} \quad \forall\, n \in \mathbb{N}.$$

(e) In (b) if $c \neq d$ then,

$$P(B_2) = P(B_2 | B_1) P(B_1) + P(B_2 | R_1) P(R_1)$$

$$= \frac{b+c}{b+r+c+d} \cdot \frac{b}{b+r} + \frac{b+d}{b+r+c+d} \cdot \frac{r}{b+r}.$$

$$= \frac{b(b+c) + r(b+d)}{(b+r+c+d)(b+r)}$$