Basic Probability, Events, and Random Variables

Matthew Seguin

1.

We are given: $\mathbb{P}[A] = 0.6$, $\mathbb{P}[B] = 0.7$, and $\mathbb{P}[C] = 0.8$.

Recall that for any event E we know $0 \leq \mathbb{P}[E] \leq 1$.

Further recall that for disjoint events F and G that $\mathbb{P}[F \cup G] = \mathbb{P}[F] + \mathbb{P}[G]$.

For any events F and G, $\mathbb{P}[F \cap G] = \mathbb{P}[F]\, \mathbb{P}[G \mid F] \leq \mathbb{P}[F]$ since the event $G \mid F$ has probability at most 1.

Alternatively $F \cap G \subset F$ so $\mathbb{P}[F \cap G] \leq \mathbb{P}[F]$ since the size of $F \cap G$ can be no larger than the size of its containter, F.

Now let us look at $\mathbb{P}[F \cup G]$ for arbitrary events F and G.

We know $F \cup G = (F \cap G^C) \cup (F \cap G) \cup (F^C \cap G)$ which is a union of disjoint events as shown below:

If $x \in F \cap G^C$ then $x \in G^C$ so $x \notin G$ and hence $x \notin F \cap G$ and $x \notin F^C \cap G$.

If $x \in F^C \cap G$ then $x \in F^C$ so $x \notin F$ and hence $x \notin F \cap G^C$ and $x \notin F \cap G$.

If $x \in F \cap G$ then $x \in F$ and $x \in G$ so $x \notin F^C$ and $x \notin G^C$.

Consequently if $x \in F \cap G$ we can say $x \notin F^C \cap G$ and $x \notin F \cap G^C$.

Since these are disjoint events we know:

$$\mathbb{P}[F \cup G] = \mathbb{P}[(F \cap G^C) \cup (F \cap G) \cup (F^C \cap G)] = \mathbb{P}[F \cap G^C] + \mathbb{P}[F \cap G] + \mathbb{P}[F^C \cap G]$$
$$= \mathbb{P}[F \cap G^C] + \mathbb{P}[F \cap G] + \mathbb{P}[F \cap G] + \mathbb{P}[F^C \cap G] - \mathbb{P}[F \cap G]$$
$$= \mathbb{P}[(F \cap G^C) \cup (F \cap G)] + \mathbb{P}[(F \cap G) \cup (F^C \cap G)] - \mathbb{P}[F \cap G]$$
$$= \mathbb{P}[F] + \mathbb{P}[G] - \mathbb{P}[F \cap G]$$

This is equivalently written as $\mathbb{P}[F \cap G] = \mathbb{P}[F] + \mathbb{P}[G] - \mathbb{P}[F \cup G]$

Furthermore we know $\mathbb{P}[F \cup G] \leq 1$ so $-\mathbb{P}[F \cup G] \geq -1$.

Then we know $\mathbb{P}[F \cap G] = \mathbb{P}[F] + \mathbb{P}[G] - \mathbb{P}[F \cup G] \geq \mathbb{P}[F] + \mathbb{P}[G] - 1$

a. Now we can apply the above results to our events A and B.

We know $\mathbb{P}[A \cap B] \leq \mathbb{P}[A] = 0.6$ and $\mathbb{P}[A \cap B] \geq \mathbb{P}[A] + \mathbb{P}[B] - 1 = 0.6 + 0.7 - 1 = 0.3$

So we have shown that $0.3 \leq \mathbb{P}[A \cap B] \leq 0.6$ $\square$

b. Now we can apply the above results to our events A, B, and C by considering the events $A \cap B$ and C.

We know $\mathbb{P}[A \cap B \cap C] = \mathbb{P}[(A \cap B) \cap C] \leq \mathbb{P}[A \cap B] \leq 0.6$.

Similarly we know $\mathbb{P}[A \cap B \cap C] = \mathbb{P}[(A \cap B) \cap C)] \geq \mathbb{P}[A \cap B] + \mathbb{P}[C] - 1 \geq 0.3 + 0.8 - 1 = 0.1$

So we have shown that $0.1 \leq \mathbb{P}[A \cap B \cap C] \leq 0.6$ $\square$

2.

We are rolling a six-sided unbiased die $n \geq 3$ times, letting R_k denote the outcome of the k-th roll and $E_{i,j}$ denote the event that roll i and roll j were the same i.e. $R_i = R_j$ means $E_{i,j} = 1$ otherwise $E_{i,j} = 0$.

Note that each roll is independent of previous ones.

First let us show that as a family $\{E_{i,j} | 1 \leq i < j \leq n\}$ is not independent:

Let $1 \leq i < j < k \leq n$ (since $n \geq 3$ such a set of (i, j, k) exists).

Assume $E_{i,j} = 1$ and $E_{j,k} = 1$, that is $R_i = R_j$ and $R_j = R_k$.

Then clearly $R_i = R_j = R_k$ so $\mathbb{P}[E_{i,k} = 1 | E_{i,j} = E_{j,k} = 1] = 1$ however clearly $\mathbb{P}[E_{i,k} = 1] \neq 1$.

So $\{E_{i,j}, E_{j,k}, E_{i,k}\}$ is not independent and since $\{E_{i,j}, E_{j,k}, E_{i,k}\} \subseteq \{E_{i,j} | 1 \leq i < j \leq n\}$ clearly $\{E_{i,j} | 1 \leq i < j \leq n\}$ is dependent as a family $\square$

Now we will show that $\{E_{i,j} | 1 \leq i < j \leq n\}$ are pairwise independent:

Recall that for events F and G if $\mathbb{P}[F \cap G] = \mathbb{P}[F]\,\mathbb{P}[G]$ then F and G are independent

(I often write $\mathbb{P}[F \cap G]$ as $\mathbb{P}[F, G]$).

First note that for $1 \leq i < j \leq n$:

$$\mathbb{P}[E_{i,j} = 1] = \mathbb{P}[R_i = R_j] = \sum_{k=1}^{6} \mathbb{P}[R_i = k | R_j = k]\mathbb{P}[R_j = k] = \sum_{k=1}^{6} \mathbb{P}[R_i = k]\left(\frac{1}{6}\right) = \frac{1}{6}\sum_{k=1}^{6}\frac{1}{6} = \frac{1}{6}$$

Let $1 \leq i < j \leq n$ and $1 \leq k < m \leq n$ with $(i, j) \neq (k, m)$.

Note that either $i \neq k$ or $j \neq m$ because of the above, meaning the following are all the possible cases:

$i \neq k$ and $j \neq m$, $i \neq k$ but $j = m$, or $j \neq m$ but $i = k$.

Now let $S = \{(x, y) \mid 1 \leq x \leq 6, 1 \leq y \leq 6\}$

(Note that there are $6^2 = 36$ distinct values in S since there are 6 choices for x and 6 choices for y for each choice of x).

- If $i \neq k$ and $j \neq m$:

All rolls are distinct and hence independent here.

Computing directly and using the independence of distinct rolls we have:

$$\mathbb{P}[E_{i,j} = 1, E_{k,m} = 1] = \mathbb{P}[R_i = R_j, R_k = R_m] = \sum_{(x,y)\in S} \mathbb{P}[R_i = x, R_k = y | R_j = x, R_m = y]\mathbb{P}[R_j = x, R_m = y]$$

$$= \sum_{(x,y)\in S} \mathbb{P}[R_i = x, R_k = y]\mathbb{P}[R_j = x]\mathbb{P}[R_m = y] = \sum_{(x,y)\in S} \mathbb{P}[R_i = x]\mathbb{P}[R_k = y]\left(\frac{1}{6}\right)^2 = \frac{1}{6^2}\sum_{(x,y)\in S}\left(\frac{1}{6}\right)^2$$

$$= \frac{1}{6^2}\sum_{p=1}^{36}\frac{1}{36} = \frac{1}{6^2} = \mathbb{P}[E_{i,j} = 1]\mathbb{P}[E_{k,m} = 1]$$

Thus if $i \neq k$ and $j \neq m$ we have shown that $E_{i,j}$ and $E_{k,m}$ are independent $\square$

- If $i \neq k$ but $j = m$:

Note that $j \neq i$ because $i < j$ and similarly $j \neq k$ because $k < m = j$.

R_i, R_j, and R_k are distinct rolls here but R_j and R_m are just the same roll.

Computing directly and using the independence of distinct rolls we have:

$$\mathbb{P}[E_{i,j} = 1, E_{k,m} = 1] = \mathbb{P}[R_i = R_j, R_k = R_m] = \mathbb{P}[R_i = R_j, R_k = R_j] = \sum_{x=1}^{6} \mathbb{P}[R_i = x, R_k = x | R_j = x]\mathbb{P}[R_j = x]$$

$$= \sum_{x=1}^{6} \mathbb{P}[R_i = x, R_k = x]\left(\frac{1}{6}\right) = \frac{1}{6}\sum_{x=1}^{6} \mathbb{P}[R_i = x]\mathbb{P}[R_k = x] = \frac{1}{6}\sum_{x=1}^{6}\left(\frac{1}{6}\right)^2 = \frac{1}{6^2} = \mathbb{P}[E_{i,j} = 1]\mathbb{P}[E_{k,m} = 1]$$

Thus if $i \neq k$ and $j = m$ we have shown that $E_{i,j}$ and $E_{k,m}$ are independent $\square$

- If $j \neq m$ but $i = k$:

Note that $i \neq j$ because $i < j$ and similarly $i \neq m$ because $i = k < m$.

R_i, R_j, and R_m are distinct rolls here but R_i and R_k are just the same roll.

Computing directly and using the independence of distinct rolls we have:

$$\mathbb{P}[E_{i,j} = 1, E_{k,m} = 1] = \mathbb{P}[R_i = R_j, R_k = R_m] = \mathbb{P}[R_j = R_i, R_m = R_i] = \sum_{x=1}^{6} \mathbb{P}[R_j = x, R_m = x | R_i = x]\mathbb{P}[R_i = x]$$

$$= \sum_{x=1}^{6} \mathbb{P}[R_j = x, R_m = x]\left(\frac{1}{6}\right) = \frac{1}{6}\sum_{x=1}^{6} \mathbb{P}[R_j = x]\mathbb{P}[R_m = x] = \frac{1}{6}\sum_{x=1}^{6}\left(\frac{1}{6}\right)^2 = \frac{1}{6^2} = \mathbb{P}[E_{i,j} = 1]\mathbb{P}[E_{k,m} = 1]$$

Thus if $j \neq m$ and $i = k$ we have shown that $E_{i,j}$ and $E_{k,m}$ are independent $\square$

So in all of the possible cases where $1 \leq i < j \leq n$ and $1 \leq k < m \leq n$ with $(i, j) \neq (k, m)$ we have shown that $E_{i,j}$ is independent from $E_{k,m}$, therefore as long as $E_{i,j}$ and $E_{k,m}$ are distinct events $E_{i,j}$ and $E_{k,m}$ are independent $\square$

3.

We are given the following probability table to start with:

	$Y=0$	$Y=1$	$Y=2$
$X=1$	1/8		
$X=2$		0	

Then we are given that $\mathbb{E}[XY] = \frac{13}{9}$ and that Y is uniform over $\{0,1,2\}$.

Let $p_{i,j} = \mathbb{P}[X=i, Y=j]$. Then the above is equivalent to:

Since this is a distribution we know $p_{1,0} + p_{2,0} + p_{1,1} + p_{2,1} + p_{1,2} + p_{2,2} = 1$.

$p_{1,0} + p_{2,0} = p_{1,1} + p_{2,1} = p_{1,2} + p_{2,2} = \frac{1}{3}$ and $p_{1,1} + 2p_{1,2} + 4p_{2,2} = \frac{13}{9}$.

We also know that $p_{2,1} = 0$ so $p_{1,1} = \frac{1}{3}$.

Similarly we know $p_{1,0} = \frac{1}{8}$ so $p_{1,0} + p_{2,0} = \frac{1}{8} + p_{2,0} = \frac{1}{3}$ and $p_{2,0} = \frac{1}{3} - \frac{1}{8} = \frac{5}{24}$.

Now notice that $p_{1,2} = \frac{1}{3} - p_{2,2}$.

Plugging in we have:

$p_{1,1} + 2p_{1,2} + 4p_{2,2} = \frac{1}{3} + 2(\frac{1}{3} - p_{2,2}) + 4p_{2,2} = 1 + 2p_{2,2} = \frac{13}{9}$.

So $2p_{2,2} = \frac{13}{9} - 1 = \frac{4}{9}$ and $p_{2,2} = \frac{2}{9}$.

Finally we know $p_{1,2} = \frac{1}{3} - p_{2,2} = \frac{1}{3} - \frac{2}{9} = \frac{1}{9}$.

Filling out the table we have:

	$Y=0$	$Y=1$	$Y=2$
$X=1$	1/8	1/3	1/9
$X=2$	5/24	0	2/9

4.

Let Y be the result of a fair coin toss, that is if the toss is tails $Y = 0$ and if the toss is heads $Y = 1$.

If we flip tails (i.e. $Y = 0$) then $X \sim Unif(0, 2)$ and if we flip heads (i.e. $Y = 1$) then $X = 1$.

Recall that if $U \sim Unif(a, b)$ then:

$$F_U(x) = \mathbb{P}[U \leq x] = \begin{cases} 0 & \text{for} \quad x < a \\ \frac{x-a}{b-a} & \text{for} \quad a \leq x \leq b \\ 1 & \text{for} \quad x > b \end{cases}$$

a.

Let $U \sim Unif(0, 2)$.

Computing directly we have:

$$\mathbb{P}[X \leq \tfrac{1}{2}] = \mathbb{P}[X \leq \tfrac{1}{2} | Y = 0]\, \mathbb{P}[Y = 0] + \mathbb{P}[X \leq \tfrac{1}{2} | Y = 1]\, \mathbb{P}[Y = 1] = \tfrac{1}{2}\Big(\mathbb{P}[X \leq \tfrac{1}{2} | Y = 0] + \mathbb{P}[X \leq \tfrac{1}{2} | Y = 1]\Big)$$

$$= \frac{1}{2}\Big(\mathbb{P}[U \leq \frac{1}{2}] + \mathbb{P}[1 \leq \frac{1}{2}]\Big) = \Big(\frac{1}{2}\Big)\Big(\frac{1/2 - 0}{2 - 0}\Big) = \frac{1}{8}$$

$$\mathbb{P}[X \leq \tfrac{3}{2}] = \mathbb{P}[X \leq \tfrac{3}{2} | Y = 0]\, \mathbb{P}[Y = 0] + \mathbb{P}[X \leq \tfrac{3}{2} | Y = 1]\, \mathbb{P}[Y = 1] = \tfrac{1}{2}\Big(\mathbb{P}[X \leq \tfrac{3}{2} | Y = 0] + \mathbb{P}[X \leq \tfrac{3}{2} | Y = 1]\Big)$$

$$= \frac{1}{2}\Big(\mathbb{P}[U \leq \frac{3}{2}] + \mathbb{P}[1 \leq \frac{3}{2}]\Big) = \Big(\frac{1}{2}\Big)\Big(\frac{3/2 - 0}{2 - 0} + 1\Big) = \frac{7}{8}$$

b.

Let $U \sim Unif(0, 2)$.

Computing directly we have:

$$F_X(x) = \mathbb{P}[X \leq x] = \mathbb{P}[X \leq x | Y = 0]\, \mathbb{P}[Y = 0] + \mathbb{P}[X \leq x | Y = 1]\, \mathbb{P}[Y = 1] = \tfrac{1}{2}\Big(\mathbb{P}[X \leq x | Y = 0] + \mathbb{P}[X \leq x | Y = 1]\Big)$$

$$= \frac{1}{2}\Big(\mathbb{P}[U \leq x] + \mathbb{P}[1 \leq x]\Big) = \frac{1}{2}\left(\begin{cases} 0 & \text{for} \quad x < 0 \\ \frac{x}{2} & \text{for} \quad 0 \leq x \leq 2 \\ 1 & \text{for} \quad x > 2 \end{cases} + \begin{cases} 0 & \text{for} \quad x < 1 \\ 1 & \text{for} \quad x \geq 1 \end{cases}\right) = \frac{1}{2}\begin{cases} 0 & \text{for} \quad x < 0 \\ \frac{x}{2} & \text{for} \quad 0 \leq x < 1 \\ \frac{x}{2} + 1 & \text{for} \quad 1 \leq x \leq 2 \\ 2 & \text{for} \quad x > 2 \end{cases}$$

$$= \begin{cases} 0 & \text{for} \quad x < 0 \\ \frac{x}{4} & \text{for} \quad 0 \leq x < 1 \\ \frac{x}{4} + \frac{1}{2} & \text{for} \quad 1 \leq x \leq 2 \\ 1 & \text{for} \quad x > 2 \end{cases}$$

c.

From the result of the previous problem we know:

$$F_X(x) = \mathbb{P}[X \leq x] = \begin{cases} 0 & \text{for} \quad x < 0 \\ \frac{x}{4} & \text{for} \quad 0 \leq x < 1 \\ \frac{x}{4} + \frac{1}{2} & \text{for} \quad 1 \leq x \leq 2 \\ 1 & \text{for} \quad x > 2 \end{cases}$$

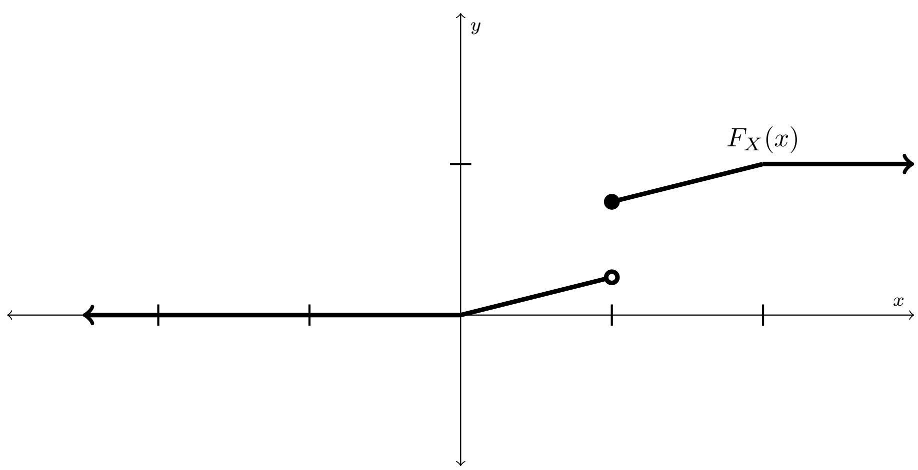
X is not a discrete random variable because it can take all of the values in $[0, 2]$ which is uncountably many, a discrete random variable must only be able to assume at most countably many values.

However X is also not a continuous random variable because $F_X(x)$ is not continuous over $\mathbb{R}$ since

$$\lim_{x \uparrow 1} F_X(x) = \lim_{x \uparrow 1} \frac{x}{4} = \frac{1}{4} \neq \frac{3}{4} = \frac{1}{4} + \frac{1}{2} = \lim_{x \downarrow 1} \frac{x}{4} + \frac{1}{2} = \lim_{x \downarrow 1} F_X(x)$$

So $\lim_{x \to 1} F_X(x)$ does not exist and hence $F_X(x)$ is not continuous at 1 and hence not continuous over $\mathbb{R}$.

We can graph this below to confirm this (tick marks are unit length apart):



5.

Let $k \in \mathbb{N}$ and let X be a random variable with finite expectation and all of its even moments defined.

That is $\mathbb{E}[X] < \infty$ and $\mathbb{E}[X^{2n}] < \infty$ for all $n \in \mathbb{N}$.

Recall Jensen's inequality that if $\phi(x)$ is a convex function and X is a random variable then $\mathbb{E}[\phi(X)] \geq \phi(\mathbb{E}[X])$.

Let $f(x) = x^{2k}$ then let $x_1,\, x_2 \in \mathbb{R}$ and $\lambda \in [0,1]$.

Recall that for $a, b \in \mathbb{R}$ that $(a+b)^2 \leq a^2 + b^2$ then we know:

$$f(\lambda x_1 + (1-\lambda)x_2) = (\lambda x_1 + (1-\lambda)x_2)^{2k} = ((\lambda x_1 + (1-\lambda)x_2)^2)^k \leq (\lambda^2 x_2^2 + (1-\lambda)^2 x_2^2))^k$$

$$= \sum_{j=0}^{k} \binom{k}{j} (\lambda^2 x_1^2)^j ((1-\lambda)^2 x_2^2)^{k-j} \leq (\lambda^2 x_1^2)^k + ((1-\lambda)^2 x_2^2)^k$$

Since every term in the sum is non-negative.

Then continuing we have:

$$f(\lambda x_1 + (1-\lambda)x_2) \leq (\lambda^2 x_1^2)^k + ((1-\lambda)^2 x_2^2)^k = \lambda^{2k} x_1^{2k} + (1-\lambda)^{2k} x_2^{2k} \leq \lambda x_1^{2k} + (1-\lambda) x_2^{2k}$$

Since $0 \leq \lambda \leq 1$ so $1 - \lambda \leq 1$. Then finally we have:

$$f(\lambda x_1 + (1-\lambda)x_2) \leq \lambda x_1^{2k} + (1-\lambda)x_2^{2k} = \lambda f(x_1) + (1-\lambda) f(x_2)$$

Showing that $f(x) = x^{2k}$ is a convex function over $\mathbb{R}$.

Therefore by Jensen's inequality we know:

$$\mathbb{E}[f(X)] = \mathbb{E}[X^{2k}] \geq (\mathbb{E}[X])^{2k} = f(\mathbb{E}[X])$$

This was true for an arbitrary $k \in \mathbb{N}$ and hence is true for all $k \in \mathbb{N}$ $\square$

6.

We are picking a point (X, Y) uniformly at random from the inside region bounded by $f(x) = 1 - x^2$, the x axis, and the y axis.

a.

We are given $f(x) = 1 - x^2$ and we want the area bounded by $f(x)$ and the x and y axes.

$f(x)$ intersects the x axis when $y = 0$ so $0 = f(x) = 1 - x^2$ implying $x = \pm 1$ but we are bounded by the y axis meaning $x > 0$ so we take $x = 1$.

Then computing the area of this region directly we get:

$$\int_0^1 1 - x^2 \, dx = x - \frac{x^3}{3}\bigg|_0^1 = 1 - \frac{1}{3} = \frac{2}{3} \quad \square$$

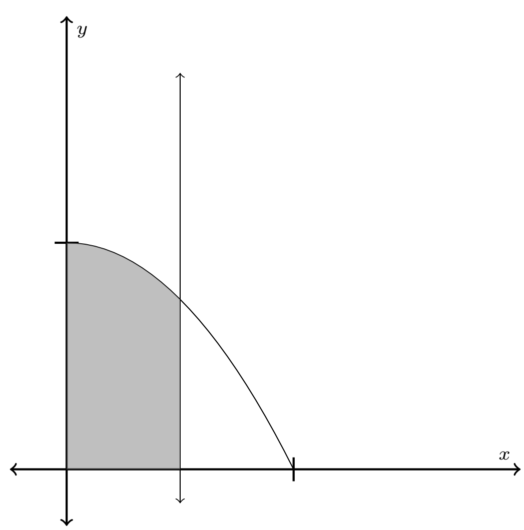
b.

Note that the closest point on the y axis from our random point (X, Y) will always be $(0, Y)$, otherwise you are increasing distance by moving up or down.

In this case the distance between these points is clearly $\sqrt{(X-0)^2 + (Y-Y)^2} = X$.

So the event "Distance to y axis $< 1/2$" is equivalent to $X < \frac{1}{2}$.

Shown in a graph this is:



Therefore to find our desired probability we can compute the area of the region under $f(x)$ where $x < \frac{1}{2}$ and divide it by the total area under $f(x)$:

$$\mathbb{P}[\text{Distance to y axis} < \frac{1}{2}] = \mathbb{P}[X < \frac{1}{2}] = \frac{1}{2/3}\int_0^1 1_{x<1/2}(1-x^2)\,dx = \frac{3}{2}\left(\int_0^{1/2} 1_{x<1/2}(1-x^2)\,dx + \int_{1/2}^1 1_{x<1/2}(1-x^2)\,dx\right)$$

$$= \frac{3}{2}\left(\int_0^{1/2} 1 - x^2\,dx + \int_{1/2}^1 0\,dx\right) = \frac{3}{2}\int_0^{1/2} 1 - x^2\,dx = \frac{3}{2}\left(x - \frac{x^3}{3}\bigg|_0^{1/2}\right) = \frac{3}{2}\left(\frac{1}{2} - \frac{1}{24}\right) = \left(\frac{3}{2}\right)\left(\frac{11}{24}\right) = \frac{11}{16}$$

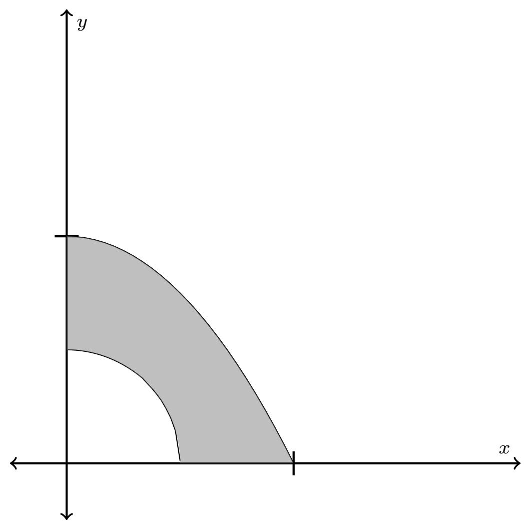
c.

The event "Distance to origin $> 1/2$" is equivalent to $\sqrt{X^2+Y^2} > \frac{1}{2}$ or equivalently:

$X^2+Y^2 > \frac{1}{4}$ and $Y^2 > \frac{1}{4} - X^2$ and $Y > \sqrt{\frac{1}{4} - X^2}$ since we know $Y > 0$ here due to the x axis bound.

Let $g(x) = \sqrt{\frac{1}{4} - x^2}$, $g(x)$ intersects the x axis when $y = 0$ so $0 = g(x) = \sqrt{\frac{1}{4} - x^2}$ implying $0 = \frac{1}{4} - x^2$ and $x = \pm\frac{1}{2}$ but we are bounded by the y axis meaning $x > 0$ so we take $x = \frac{1}{2}$.

Shown in a graph this is:



Then to find our desired probability we can first find the complement probability by computing the area of the region under $g(x)$ and divide it by the total area under $f(x)$:

$$\mathbb{P}[\text{Distance to origin} > \frac{1}{2}] = \mathbb{P}[\sqrt{X^2+Y^2} > \frac{1}{2}] = 1 - \mathbb{P}[\sqrt{X^2+Y^2} \leq \frac{1}{2}] = 1 - \frac{1}{2/3}\int_0^{1/2} \sqrt{\frac{1}{4} - x^2}\, dx$$

$$= 1 - \frac{3}{2}\int_0^{1/2} \sqrt{\frac{1}{4} - x^2}\, dx$$

Now note that the area under $\sqrt{\frac{1}{4} - x^2}$ from $x = 0$ to $x = 1/2$ is just a quarter of the area of the circle centered at the origin with radius $1/2$. Therefore:

$$\mathbb{P}[\text{Distance to origin} > \frac{1}{2}] = 1 - \frac{3}{2}\int_0^{1/2} \sqrt{\frac{1}{4} - x^2}\, dx = 1 - \frac{3}{2}\Big(\frac{\pi(1/2)^2}{4}\Big) = 1 - \frac{3\pi}{32} \quad \square$$

d.

First let us find the CDF of X by computing the area under $f(x)$ where $X \leq x$ and dividing by the total area under $f(x)$: Note that for $x \leq 0$ we know $F_X(x) = \mathbb{P}[X \leq x] = 0$ since we only consider points bounded by the y axis (i.e. $x > 0$). Also note that for $x \geq 1$ we know $F_X(x) = \mathbb{P}[X \leq x] = 1$ since we only consider points bounded by $f(x)$ and the x axis which meet at $x = 1$ and hence any point chosen has x value less than 1.

Now let $0 < x < 1$, computing directly we have:

$$F_X(x) = \mathbb{P}[X \leq x] = \frac{1}{2/3}\int_0^x 1 - x^2 \, dx = \frac{3}{2}\Big(x - \frac{x^3}{3}\Big|_0^x\Big) = \frac{3}{2}x - \frac{x^3}{2}$$

So we have the CDF as:

$$F_X(x) = \mathbb{P}[X \leq x] = \begin{cases} 0 & \text{for} \quad x \leq 0 \\ \frac{3}{2}x - \frac{x^3}{2} & \text{for} \quad 0 < x < 1 \\ 1 & \text{for} \quad x \geq 1 \end{cases}$$

First note that this piecewise function is differentiable in x over $(0, 1)$ since it is a polynomial of x for $x \in (0, 1)$. That was the CDF of X so to find the PDF we can take the derivative with respect to x in the range of values X takes and make the density 0 when we are out of range of possible points for X.

$$f_X(x) = \begin{cases} 0 & \text{for} \quad x \leq 0 \\ \dfrac{\partial}{\partial x}\dfrac{3}{2}x - \dfrac{x^3}{2} & \text{for} \quad 0 < x < 1 \\ 0 & \text{for} \quad x \geq 1 \end{cases} = \begin{cases} 0 & \text{for} \quad x \leq 0 \\ \dfrac{3}{2}\Big(1 - x^2\Big) & \text{for} \quad 0 < x < 1 \\ 0 & \text{for} \quad x \geq 1 \end{cases}$$

e.

First let us find the CDF of Y by computing the area under $f(x)$ where $Y \leq y$ and dividing by the total area under $f(x)$:

Note that for $y \leq 0$ we know $F_Y(y) = \mathbb{P}[Y \leq y] = 0$ since we only consider points bounded by the x axis (i.e. $y > 0$).

Also note that for $y \geq 1$ we know $F_Y(y) = \mathbb{P}[Y \leq y] = 1$ since we only consider points bounded by $f(x)$ and the y axis which meet at $y = 1$ and hence any point chosen has y value less than 1.

Now let $0 < y < 1$, computing directly we have:

First we will solve for $x(y)$: $y = f(x) = 1 - x^2$ so $1 - y = x^2$ and $x = \sqrt{1-y}$ since $x > 0$.

$$F_Y(y) = \mathbb{P}[Y \leq y] = \frac{1}{2/3} \int_0^y \sqrt{1-y}\, dy$$

Let $u = 1 - y$ then $\frac{du}{dy} = -1$ and $u(0) = 1$, $u(y) = 1 - y$.

$$F_Y(y) = \frac{1}{2/3} \int_0^y \sqrt{1-y}\, dy = -\frac{3}{2} \int_1^{1-y} u^{1/2}\, du = \frac{3}{2} \Big(\frac{2}{3} u^{3/2} \Big|_{1-y}^{1} \Big) = 1 - (1-y)^{3/2}$$

So we have the CDF as:

$$F_Y(y) = \mathbb{P}[Y \leq y] = \begin{cases} 0 & \text{for} \quad y \leq 0 \\ 1 - (1-y)^{3/2} & \text{for} \quad 0 < y < 1 \\ 1 & \text{for} \quad y \geq 1 \end{cases}$$

First note that this piecewise function is differentiable in y over $(0, 1)$ since it is a polynomial of y multiplied by $\sqrt{1-y}$ and added to a constant (all of which are differentiable functions) for $y \in (0, 1)$.

That was the CDF of Y so to find the PDF we can take the derivative with respect to y in the range of values Y takes and make the density 0 when we are out of range of possible points for Y.

$$f_Y(y) = \begin{cases} 0 & \text{for} \quad y \leq 0 \\ \dfrac{\partial}{\partial y} 1 - (1-y)^{3/2} & \text{for} \quad 0 < y < 1 \\ 0 & \text{for} \quad y \geq 1 \end{cases} = \begin{cases} 0 & \text{for} \quad y \leq 0 \\ \frac{3}{2}\sqrt{1-y} & \text{for} \quad 0 < y < 1 \\ 0 & \text{for} \quad y \geq 1 \end{cases}$$

f.

X and Y are dependent because if we know $X = x$ then Y must lie between 0 and $f(x) = 1 - x^2$ which clearly relies on x.

Formally this is written as $Y|X = x \in (0, f(x)) = (0, 1 - x^2)$.

As an example:

Since all points between $f(x)$ and both the x axis and y axis are possible.

If $X = 1/3$ then Y can take any value in $(0, 1 - (1/3)^2) = (0, 8/9)$.

While if $X = 2/3$ then Y can take any value in $(0, 1 - (2/3)^2) = (0, 5/9)$.

So the range of Y has changed based upon what value X assumes implying that the distribution of Y changed based upon what value X assumes showing that Y is dependent upon X and hence X and Y are not independent $\square$

7.

a.

Let $A_1, A_2, ..., A_n$ be events and their corresponding indicators be $I(A_1), I(A_2), ..., I(A_n)$.

Then we know that:

$$x \in \bigcup_{i=1}^{n} A_i \text{ if and only if } x \in A_i \text{ for some } i \in \{1, 2, ..., n\}$$

$$\text{Therefore } I\Big(\bigcup_{i=1}^{n} A_i\Big) = 1 \text{ if and only if } I(A_i) = 1 \text{ for some } i \in \{1, 2, ..., n\}$$

Similarly:

$$x \notin \bigcup_{i=1}^{n} A_i \text{ if and only if } x \notin A_i \text{ for all } i \in \{1, 2, ..., n\}$$

$$\text{Therefore } I\Big(\bigcup_{i=1}^{n} A_i\Big) = 0 \text{ if and only if } I(A_i) = 0 \text{ for all } i \in \{1, 2, ..., n\}$$

Since $I(A_i) \in \{0, 1\}$ for all $i \in \{1, 2, ..., n\}$ we know that:

$\max_{1 \leq i \leq n} I(A_i) = 1$ if and only if $I(A_i) = 1$ for some $i \in \{1, 2, ..., n\}$

$\max_{1 \leq i \leq n} I(A_i) = 0$ if and only if $I(A_i) = 0$ for all $i \in \{1, 2, ..., n\}$

Equivalently this is:

$$I\Big(\bigcup_{i=1}^{n} A_i\Big) = 1 \text{ if and only if } \max_{1 \leq i \leq n} I(A_i) = 1$$

$$I\Big(\bigcup_{i=1}^{n} A_i\Big) = 0 \text{ if and only if } \max_{1 \leq i \leq n} I(A_i) = 0$$

Which means:

$$I\Big(\bigcup_{i=1}^{n} A_i\Big) = \max_{1 \leq i \leq n} I(A_i) \ \square$$

b.

From the previous part we have the result:

$$I\Big(\bigcup_{i=1}^{n} A_i\Big) = \max_{1\leq i\leq n} I(A_i)$$

Note that if $\max_{1\leq i\leq n} I(A_i) = 0$ then:

$$\sum_{i=1}^{n} I(A_i) = 0$$

Also if $\max_{1\leq i\leq n} I(A_i) = 1$ then:

$$\sum_{i=1}^{n} I(A_i) \geq 1$$

Therefore we know:

$$\max_{1\leq i\leq n} I(A_i) \leq \sum_{i=1}^{n} I(A_i)$$

Then recall that if $X \leq Y$ then $\mathbb{E}[X] \leq \mathbb{E}[Y]$ to give us the result:

$$\mathbb{P}[\bigcup_{i=1}^{n} A_i] = \mathbb{E}[I\Big(\bigcup_{i=1}^{n} A_i\Big)] = \mathbb{E}[\max_{1\leq i\leq n} I(A_i)] \leq \mathbb{E}[\sum_{i=1}^{n} I(A_i)] = \sum_{i=1}^{n} \mathbb{E}[I(A_i)] = \sum_{i=1}^{n} \mathbb{P}[A_i] \ \square$$

c.

Note that if $x \in A^c$ if and only if $x \notin A$. Similarly $x \notin A^c$ if and only if $x \in A$.

Therefore $I(A^c) = 1$ if and only if $I(A) = 0$. Or equivalently $I(A^c) = 1$ if and only if $1 - I(A) = 1$.

Similarly $I(A^c) = 0$ if and only if $I(A) = 1$. Or equivalently $I(A^c) = 0$ if and only if $1 - I(A) = 0$.

This gives us the result:

$I(A^c) = 1 - I(A)$ $\square$

d.

Let $A_1, A_2, ..., A_n$ be events and their corresponding indicators be $I(A_1), I(A_2), ..., I(A_n)$.

Then we know that:

$$x \in \bigcap_{i=1}^{n} A_i \text{ if and only if } x \in A_i \text{ for all } i \in \{1, 2, ..., n\}$$

$$\text{Therefore } I\Big(\bigcap_{i=1}^{n} A_i\Big) = 1 \text{ if and only if } I(A_i) = 1 \text{ for all } i \in \{1, 2, ..., n\}$$

Similarly:

$$x \notin \bigcap_{i=1}^{n} A_i \text{ if and only if } x \notin A_i \text{ for some } i \in \{1, 2, ..., n\}$$

$$\text{Therefore } I\Big(\bigcap_{i=1}^{n} A_i\Big) = 0 \text{ if and only if } I(A_i) = 0 \text{ for some } i \in \{1, 2, ..., n\}$$

Since $I(A_i) \in \{0, 1\}$ for all $i \in \{1, 2, ..., n\}$ we know that:

$$\prod_{i=1}^{n} I(A_i) = 1 \text{ if and only if } I(A_i) = 1 \text{ for all } i \in \{1, 2, ..., n\}$$

$$\prod_{i=1}^{n} I(A_i) = 0 \text{ if and only if } I(A_i) = 0 \text{ for some } i \in \{1, 2, ..., n\}$$

Equivalently this is:

$$I\Big(\bigcap_{i=1}^{n} A_i\Big) = 1 \text{ if and only if } \prod_{i=1}^{n} I(A_i) = 1$$

$$I\Big(\bigcap_{i=1}^{n} A_i\Big) = 0 \text{ if and only if } \prod_{i=1}^{n} I(A_i) = 0$$

Which means:

$$I\Big(\bigcap_{i=1}^{n} A_i\Big) = \prod_{i=1}^{n} I(A_i) \ \square$$

e.

Base case ($n=1$):

$$\mathbb{P}[\bigcup_{i=1}^{n} A_i] = \mathbb{P}[\bigcup_{i=1}^{1} A_i] = \mathbb{P}[A_1] = (-1)^{1-1}\mathbb{P}[A_1] = \sum_{k=1}^{1}(-1)^{k-1}\mathbb{P}[A_1 \cap A_1 \cap ... \cap A_1]$$

$$= \sum_{k=1}^{1}(-1)^{k-1} \sum_{1 \le i_1 \le i_2 \le ... \le i_k \le 1} \mathbb{P}[A_{i_1} \cap A_{i_2} \cap ... \cap A_{i_k}] = \sum_{k=1}^{n}(-1)^{k-1} \sum_{1 \le i_1 \le i_2 \le ... \le i_k \le n} \mathbb{P}[A_{i_1} \cap A_{i_2} \cap ... \cap A_{i_k}] \ \square$$

Inductive step (n implies $n+1$):

Assume that $\mathbb{P}[\bigcup_{i=1}^{n} B_i] = \sum_{k=1}^{n}(-1)^{k-1} \sum_{1 \le i_1 \le i_2 \le ... \le i_k \le n} \mathbb{P}[B_{i_1} \cap B_{i_2} \cap ... \cap B_{i_k}]$ for arbitrary $B_1, B_2, ..., B_n$

Then:

$$1 - \mathbb{P}[\bigcup_{i=1}^{n+1} A_i] = \mathbb{P}[\Big(\bigcup_{i=1}^{n+1} A_i\Big)^c] = \mathbb{P}[A_{n+1}^c \cap \Big(\bigcup_{i=1}^{n} A_i\Big)^c] = \mathbb{E}[I\Big(A_{n+1}^c \cap \Big(\bigcup_{i=1}^{n} A_i\Big)^c\Big)] = \mathbb{E}[I\Big(A_{n+1}^c\Big)I\Big(\Big(\bigcup_{i=1}^{n} A_i\Big)^c\Big)]$$

$$= \mathbb{E}[\Big(1 - I\Big(A_{n+1}\Big)\Big)\Big(1 - I\Big(\bigcup_{i=1}^{n} A_i\Big)\Big)] = \mathbb{E}[1 - I\Big(\bigcup_{i=1}^{n} A_i\Big) - I\Big(A_{n+1}\Big) + I\Big(A_{n+1}\Big)I\Big(\bigcup_{i=1}^{n} A_i\Big)]$$

$$= 1 - \mathbb{P}[\bigcup_{i=1}^{n} A_i] - \mathbb{P}[A_{n+1}] + \mathbb{E}[I\Big(A_{n+1} \cap \bigcup_{i=1}^{n} A_i\Big)] = 1 - \mathbb{P}[\bigcup_{i=1}^{n} A_i] - \mathbb{P}[A_{n+1}] + \mathbb{P}[\bigcup_{i=1}^{n}(A_i \cap A_{n+1})]$$

Which is equivalent to:

$$\mathbb{P}[\bigcup_{i=1}^{n+1} A_i] = \mathbb{P}[\bigcup_{i=1}^{n} A_i] + \mathbb{P}[A_{n+1}] - \mathbb{P}[\bigcup_{i=1}^{n}(A_i \cap A_{n+1})]$$

$$= \sum_{k=1}^{n}(-1)^{k-1} \sum_{1 \le i_1 \le i_2 \le ... \le i_k \le n} \mathbb{P}[A_{i_1} \cap A_{i_2} \cap ... \cap A_{i_k}] + \mathbb{P}[A_{n+1}] - \sum_{k=1}^{n}(-1)^{k-1} \sum_{1 \le i_1 \le i_2 \le ... \le i_k \le n} \mathbb{P}[A_{i_1} \cap A_{i_2} \cap ... \cap A_{i_k} \cap A_{n+1}]$$

$$= \sum_{i=1}^{n+1} \mathbb{P}[A_k] + \sum_{k=2}^{n}(-1)^{k-1} \sum_{1 \le i_1 \le i_2 \le ... \le i_k \le n} \mathbb{P}[A_{i_1} \cap A_{i_2} \cap ... \cap A_{i_k}] - \sum_{k=1}^{n}(-1)^{k-1} \sum_{1 \le i_1 \le i_2 \le ... \le i_k \le n} \mathbb{P}[A_{i_1} \cap A_{i_2} \cap ... \cap A_{i_k} \cap A_{n+1}]$$

$$= \sum_{i=1}^{n+1} \mathbb{P}[A_k] + \sum_{k=2}^{n+1}(-1)^{k-1} \sum_{1 \le i_1 \le i_2 \le ... \le i_k \le n+1} \mathbb{P}[A_{i_1} \cap A_{i_2} \cap ... \cap A_{i_k}]$$

$$= \sum_{k=1}^{n+1}(-1)^{k-1} \sum_{1 \le i_1 \le i_2 \le ... \le i_k \le n+1} \mathbb{P}[A_{i_1} \cap A_{i_2} \cap ... \cap A_{i_k}]$$

Therefore by induction:

Therefore $\mathbb{P}[\bigcup_{i=1}^{n} B_i] = \sum_{k=1}^{n}(-1)^{k-1} \sum_{1 \le i_1 \le i_2 \le ... \le i_k \le n} \mathbb{P}[B_{i_1} \cap B_{i_2} \cap ... \cap B_{i_k}]$ for all $n \in \mathbb{N}$ and arbitrary events $B_1, B_2, ..., B_n$ $\square$

8.

There are R red balls and $N - R$ white balls $(0 < R < N)$.

Let R_i denote the event that the ith ball drawn without replacement is red.

a.

Consider a sample of size N, that is we will get all of the balls in our sample.

We can change the order of the elements in this sample to get any of the other possible samples.

There are $N!$ such samples since all we need to do is choose where the red balls go and the blue balls will be automatically assigned. All of these samples are equally likely by the properties of the hypergeometric distribution.

Therefore $\mathbb{P}[R_i] = \mathbb{P}[\text{red ball in spot i after randomizing the order of the sample}]$

Fix one of the R red balls is in spot i, there are R ways to do this. Then there are $(N-1)!$ ways to assign the remaining balls, hence there are $R(N-1)!$ permutations in which a red ball is in spot i.

Therefore:

$$\mathbb{P}[R_i] = \frac{R(N-1)!}{N!} = \frac{R}{N} \quad \square$$

This also works if we consider a sample of size $k \leq N$ because we can just consider the set of all samples of size N where we got the same first k draws as in our sample of size k.

b.

From the previous problem we know $\mathbb{P}[R_i] = \frac{R}{N}$ for all $i \in \{1, ..., N\}$.

Therefore:

$$\mathbb{P}[R_i|R_j] = \frac{\mathbb{P}[R_i, R_j]}{\mathbb{P}[R_j]} = \frac{\mathbb{P}[R_j, R_i]}{\frac{R}{N}} = \frac{\mathbb{P}[R_j, R_i]}{\mathbb{P}[R_i]} = \mathbb{P}[R_j|R_i] \quad \square$$

c.

First note that the following events means at least one of draws $k+1$ to n is a red:

We can calculate this probability by taking the complement (which says that draws $k+1$ to n are all not red).

$$\mathbb{P}[\bigcup_{i=k+1}^{n} R_i] = 1 - \mathbb{P}[\bigcap_{i=k+1}^{n} R_i^c]$$

First choose $n-k$ white balls to assign to spots $k+1$ to n, there are $(N-R)(N-R-1)...(N-R-n+k+1) = \frac{(N-R)!}{(N-R-n+k)!}$ ways to do this. Then we have $N-n+k$ balls left to assign in which we can do $(N-n+k)!$ ways.

So there are $\frac{(N-R)!(N-n+k)!}{(N-R-n+k)!}$ permutations in which draws $k+1$ to n are all not red, therefore:

$$\mathbb{P}[\bigcap_{i=k+1}^{n} R_i^c] = \frac{\frac{(N-R)!(N-n+k)!}{(N-R-n+k)!}}{N!} = \frac{(N-R)!(N-n+k)!}{N!(N-R-n+k)!} = \frac{\binom{N-R}{n-k}}{\binom{N}{n-k}}$$

Implying:

$$\mathbb{P}[\bigcup_{i=k+1}^{n} R_i] = 1 - \mathbb{P}[\bigcap_{i=k+1}^{n} R_i^c] = 1 - \frac{\binom{N-R}{n-k}}{\binom{N}{n-k}} = \frac{\binom{N}{n-k} - \binom{N-R}{n-k}}{\binom{N}{n-k}}$$

Then we can compute:

$$\mathbb{P}[\bigcup_{i=k+1}^{n} R_i \mid R_k] = 1 - \mathbb{P}[\bigcap_{i=k+1}^{n} R_i^c \mid R_k]$$

Assume we already know R_k, that is draw k is a red ball. There are $(N-1)!$ total ways to assign the remaining balls after we know this, so here we will divide by $(N-1)!$ rather than $N!$ to calculate the probability.

Now choose $n-k$ white balls to assign to spots $k+1$ to n, there are $(N-R)(N-R-1)...(N-R-n+k+1) = \frac{(N-R)!}{(N-R-n+k)!}$ ways to do this. Then we have $N-n+k-1$ balls left to assign in which we can do $(N-n+k-1)!$ ways.

So there are $\frac{(N-R)!(N-n+k-1)!}{(N-R-n+k)!}$ permutations in which draws $k+1$ to n are all not red given we already know draw k is red, therefore:

$$\mathbb{P}[\bigcap_{i=k+1}^{n} R_i^c \mid R_k] = \frac{\frac{(N-R)!(N-n+k-1)!}{(N-R-n+k)!}}{(N-1)!} = \frac{(N-R)!(N-n+k-1)!}{(N-1)!(N-R-n+k)!} = \frac{\binom{N-R}{n-k}}{\binom{N-1}{n-k}}$$

Implying:

$$\mathbb{P}[\bigcup_{i=k+1}^{n} R_i \mid R_k] = 1 - \mathbb{P}[\bigcap_{i=k+1}^{n} R_i^c \mid R_k] = 1 - \frac{\binom{N-R}{n-k}}{\binom{N-1}{n-k}} = \frac{\binom{N-1}{n-k} - \binom{N-R}{n-k}}{\binom{N-1}{n-k}}$$

Now recall from part a we know $\mathbb{P}[R_k] = \frac{R}{N}$, then we have the result:

$$\mathbb{P}[R_k \mid \bigcup_{i=k+1}^{n} R_i] = \frac{\mathbb{P}[\bigcup_{i=k+1}^{n} R_i \mid R_k]\mathbb{P}[R_k]}{\mathbb{P}[\bigcup_{i=k+1}^{n} R_i]}$$

$$= \frac{\left(\frac{\binom{N-1}{n-k}-\binom{N-R}{n-k}}{\binom{N-1}{n-k}}\right)\left(\frac{R}{N}\right)}{\frac{\binom{N}{n-k}-\binom{N-R}{n-k}}{\binom{N}{n-k}}} = \left(\frac{R}{N}\right)\left(\frac{\binom{N}{n-k}}{\binom{N-1}{n-k}}\right)\left(\frac{\binom{N-1}{n-k} - \binom{N-R}{n-k}}{\binom{N}{n-k} - \binom{N-R}{n-k}}\right)$$

$$= \left(\frac{R}{N}\right)\left(\frac{\frac{N!}{(N-n+k)!(n-k)!}}{\frac{(N-1)!}{(N-n+k-1)!(n-k)!}}\right)\left(\frac{\binom{N-1}{n-k} - \binom{N-R}{n-k}}{\binom{N}{n-k} - \binom{N-R}{n-k}}\right) = \left(\frac{R}{N}\right)\left(\frac{N!(N-n+k-1)!(n-k)!}{(N-n+k)!(n-k)!(N-1)!}\right)\left(\frac{\binom{N-1}{n-k} - \binom{N-R}{n-k}}{\binom{N}{n-k} - \binom{N-R}{n-k}}\right)$$

$$= \left(\frac{R}{N}\right)\left(\frac{N}{N-n+k}\right)\left(\frac{\binom{N-1}{n-k} - \binom{N-R}{n-k}}{\binom{N}{n-k} - \binom{N-R}{n-k}}\right) = \left(\frac{R}{N-n+k}\right)\left(\frac{\binom{N-1}{n-k} - \binom{N-R}{n-k}}{\binom{N}{n-k} - \binom{N-R}{n-k}}\right) \quad \square$$

d.

Let $0 < R < N$, then:

$$X_{R,N} = min\{k \geq 1 \ : \ \sum_{i=1}^{k} I(R_i) \geq 1 \text{ when there are } R \text{ red balls and } N \text{ total balls to start}\}$$

$$= min\{k \geq 1 \ : \ R_k \text{ when there are } R \text{ red balls and } N \text{ total balls to start}\}$$

That is X is the number of the first draw where we get a red ball from an urn with R red balls and N total balls to start.

We can break this down into an iterative process.

If we didn't draw a red ball on the first draw then what we have left is R red balls and $N-1$ total balls so

$$X_{R,N}| \text{ first draw isn't red} \sim 1 + X_{R,N-1}$$

If we did draw a red ball on the first draw then we are just done so $X_{R,N}| \text{ first draw is red} = 1$

Let $\mu_{R,N} = \mathbb{E}[X_{R,N}]$, then from above we can see:

$$\mu_{R,N} = \mathbb{P}[\text{red}] + \mathbb{P}[\text{not red}](1 + \mu_{R,N-1}) = \mathbb{P}[\text{red}] + \mathbb{P}[\text{not red}] + \mu_{R,N-1}\mathbb{P}[\text{not red}]$$

$$= 1 + \mu_{R,N-1}\mathbb{P}[\text{not red}] = 1 + \frac{N-R}{N}\mu_{R,N-1}$$

Let $r, w > 0$, then:

Starting at the trivial case we have $\mu_{r,r} = 1$ since we must draw a red ball if they are all red.

Then $\mu_{r,r+1} = 1 + \frac{1}{r+1}\mu_{r,r} = \frac{r+2}{r+1}$.

Then $\mu_{r,r+2} = 1 + \frac{2}{r+2}\mu_{r,r+1} = \frac{r+3}{r+1}$.

We can already suspect a pattern here, namely we expect to see $\mu_{r,r+w} = \frac{r+w+1}{r+1}$.

We have already seen the base case is true ($\mu_{r,r+0} = \frac{r+0+1}{r+1} = 1$).

Now assume that $\mu_{r,r+w} = \frac{r+w+1}{r+1}$, then:

$$\mu_{r,r+w+1} = 1 + \frac{w+1}{r+w+1}\mu_{r,r+w} = 1 + \Big(\frac{w+1}{r+w+1}\Big)\Big(\frac{r+w+1}{r+1}\Big) = 1 + \frac{w+1}{r+1} = \frac{r+w+2}{r+1}$$

Thus we have completed the inductive step and therefore if $r, w > 0$ then $\mu_{r,r+w} = \frac{r+w+1}{r+1}$ for all $w \geq 0$.

Finally if $0 < R < N$ and there are R red balls and $N - R$ white balls (i.e. N total balls) in an urn we have the result:

$$\mathbb{E}[X_{R,N}] = \mu_{R,N} = \mu_{R,R+(N-R)} = \frac{R + (N-R) + 1}{R+1} = \frac{N+1}{R+1} \ \square$$

e.

Consider an urn with N_i balls of color i for each $i \in \{1, 2, ..., k\}$ where $N_1 + N_2 + ... + N_k = N$.

We are looking for how we can draw r_i balls of color i for each $i \in \{1, 2, ..., k\}$.

As before the probability of every sample of size n is equally likely.

There are $N(N-1)...(N-n+1) = \frac{N!}{(N-n)!}$ samples of size n.

Consider a sequence of draws in which we get r_1 balls of color 1 first then r_2 balls of color 2 and so on until we get r_k balls of color k.

The number of ways to get such a sequence is given by:

$$\Big((N_1)(N_1-1)...(N_1-r_1+1)\Big)\Big((N_2)(N_2-1)...(N_2-r_2+1)\Big)...\Big((N_k)(N_k-1)...(N_k-r_k+1)\Big)$$

$$= \Big(\frac{N_1!}{(N_1-r_1)!}\Big)\Big(\frac{N_2!}{(N_2-r_2)!}\Big)...\Big(\frac{N_k!}{(N_k-r_k)!}\Big)$$

Therefore the probability of such a sequence is:

$$\frac{\Big(\frac{N_1!}{(N_1-r_1)!}\Big)\Big(\frac{N_2!}{(N_2-r_2)!}\Big)...\Big(\frac{N_k!}{(N_k-r_k)!}\Big)}{\frac{N!}{(N-n)!}}$$

There are $\binom{n}{r_1}\binom{n-r_1}{r_2}...\binom{n-r_1-r_2-...-r_{k-2}}{r_{k-1}}$ samples where we get r_i balls of color i for each $i \in \{1, 2, ..., k\}$ since we just need to choose where all of the balls of the first $k-1$ colors go and we will know where the balls of the last color go.

Therefore:

$$\mathbb{P}[\text{Draw } r_i \text{ balls of color } i \text{ for each } i \in \{1, 2, ..., k\}]$$

$$= \frac{\Big(\frac{N_1!}{(N_1-r_1)!}\Big)\Big(\frac{N_2!}{(N_2-r_2)!}\Big)...\Big(\frac{N_k!}{(N_k-r_k)!}\Big)}{\frac{N!}{(N-n)!}}\binom{n}{r_1}\binom{n-r_1}{r_2}...\binom{n-r_1-r_2-...-r_{k-2}}{r_{k-1}}$$

$$= \frac{\Big(\frac{N_1!}{(N_1-r_1)!r_1!}\Big)\Big(\frac{N_2!}{(N_2-r_2)!}\Big)...\Big(\frac{N_k!}{(N_k-r_k)!}\Big)}{(n-r_1)!\Big(\frac{N!}{(N-n)!n!}\Big)}\binom{n-r_1}{r_2}\binom{n-r_1-r_2}{r_3}...\binom{n-r_1-r_2-...-r_{k-2}}{r_{k-1}}$$

$$= \frac{\binom{N_1}{r_1}\Big(\frac{N_2!}{(N_2-r_2)!r_2!}\Big)...\Big(\frac{N_k!}{(N_k-r_k)!}\Big)}{(n-r_1-r_2)!\binom{N}{n}}\binom{n-r_1-r_2}{r_3}\binom{n-r_1-r_2-r_3}{r_4}...\binom{n-r_1-r_2-...-r_{k-2}}{r_{k-1}}$$

$$= \frac{\binom{N_1}{r_1}\binom{N_2}{r_2}\Big(\frac{N_3!}{(N_3-r_3)!}\Big)...\Big(\frac{N_k!}{(N_k-r_k)!}\Big)}{(n-r_1-r_2)!\binom{N}{n}}\binom{n-r_1-r_2}{r_3}\binom{n-r_1-r_2-r_3}{r_4}...\binom{n-r_1-r_2-...-r_{k-2}}{r_{k-1}}$$

Notice: each time we change $\binom{n-r_1-...-r_{j-1}}{r_j}$ to a ratio of factorials and bring those terms into the fraction, $\frac{N_j!}{(N_j-r_j)!}$ changes to $\binom{N_j}{r_j}$ and the term in the denominator to change from $(n-r_1-...-r_{j-1})!$ to $(n-r_1-...-r_j)!$ since we are multiplying by $\frac{(n-r_1-...-r_{j-1})!}{r_j!(n-r_1-...-r_{j-1}-r_j)!}$ which is equivalent to dividing the numerator by $r_j!$ and multiplying the denominator by $\frac{(n-r_1-...-r_{j-1}-r_j)!}{(n-r_1-...-r_{j-1})!}$

Continuing this pattern until we have reached the $\binom{n-r_1-...-r_{k-2}}{r_{k-1}}$ term, we proceed:

$$\mathbb{P}[\text{Draw } r_i \text{ balls of color } i \text{ for each } i \in \{1,2,...,k\}]$$

$$= \frac{\binom{N_1}{r_1}\binom{N_2}{r_2}\left(\frac{N_3!}{(N_3-r_3)!}\right)...\left(\frac{N_k!}{(N_k-r_k)!}\right)}{(n-r_1-r_2)!\binom{N}{n}}\binom{n-r_1-r_2}{r_3}\binom{n-r_1-r_2-r_3}{r_4}...\binom{n-r_1-r_2-...-r_{k-2}}{r_{k-1}}$$

$$\vdots$$

$$= \frac{\binom{N_1}{r_1}\binom{N_2}{r_2}...\binom{N_{k-2}}{r_{k-2}}\left(\frac{N_{k-1}!}{(N_{k-1}-r_{k-1})!}\right)\left(\frac{N_k!}{(N_k-r_k)!}\right)}{(n-r_1-r_2-...-r_{k-2})!\binom{N}{n}}\binom{n-r_1-r_2-...-r_{k-2}}{r_{k-1}}$$

$$= \frac{\binom{N_1}{r_1}\binom{N_2}{r_2}...\binom{N_{k-2}}{r_{k-2}}\left(\frac{N_{k-1}!}{(N_{k-1}-r_{k-1})!r_{k-1}!}\right)\left(\frac{N_k!}{(N_k-r_k)!}\right)}{(n-r_1-r_2-...-r_{k-2}-r_{k-1})!\binom{N}{n}} = \frac{\binom{N_1}{r_1}\binom{N_2}{r_2}...\binom{N_{k-2}}{r_{k-2}}\binom{N_{k-1}}{r_{k-1}}\left(\frac{N_k!}{(N_k-r_k)!}\right)}{r_k!\binom{N}{n}}$$

$$= \frac{\binom{N_1}{r_1}\binom{N_2}{r_2}...\binom{N_{k-2}}{r_{k-2}}\binom{N_{k-1}}{r_{k-1}}\left(\frac{N_k!}{(N_k-r_k)!r_k!}\right)}{\binom{N}{n}} = \frac{\binom{N_1}{r_1}\binom{N_2}{r_2}...\binom{N_{k-2}}{r_{k-2}}\binom{N_{k-1}}{r_{k-1}}\binom{N_k}{r_k}}{\binom{N}{n}} \quad \square$$