Problem of the Month
Solution to Problem 3: December 2022

(a) Suppose there are r red marbles and b blue marbles and set $n = r + b$. Upon removing the n marbles from the bag in random order, we can record the order in which they emerged from the bag by a sequence of R's and B's. For example, if the first two marbles were red and the third was blue, then the sequence would begin with RRB. If the fourth was red, then the sequence would continue $RRBR$, and so on.

An important observation in this part and later parts of the solution is that every possible sequence of n characters, r of which are R and b of which are B, is equally likely to occur through this process. To make this observation, we first imagine labelling the red marbles by R_1, R_2, and so on through to R_r, and the blue marbles by B_1, B_2, and so on through to B_b. With this labelling, we have made it so that the n marbles are distinct. There are exactly $n!$ orders in which the marbles can be removed, and each of them is equally likely. If we now imagine listing these $n!$ possible orderings of $R_1, \ldots, R_r, B_1, \ldots, B_b$ and removing the subscripts, each possible sequence of r R's and b B's will occur in the list exactly $(r!)(b!)$ times. Therefore, if we remove the (unindexed) marbles and write down the sequence of R's and B's, each possible sequence will occur with probability exactly $\dfrac{r!b!}{n!} = \dfrac{1}{\binom{n}{r}}$. Hence, every sequence occurs with the same probability. Notice that $\binom{n}{r}$ is equal to the number of sequences of r R's and b B's since any such sequence is completely determined by selecting r of the n possible positions and placing R in them.

Therefore, the probability that the red marbles are removed first is equal to the number of sequences of r R's and b B's with the property that there is at least one B to the right of the rightmost R, divided by $\binom{n}{r}$.

Since there are only two colours in this part of the problem, the statement "there is at least one B to the right of the rightmost R" is equivalent to "the final letter in the sequence is B". Put in a different way, the red marbles are completely removed first if and only if a blue marble is removed last. Therefore, the probability that we seek is simply the number of sequences of R's and B's that end in B, divided by $\binom{n}{r}$. The number of such sequences is equal to $\binom{n-1}{r}$ since, to construct all such sequences, we can place one B at the end, and then place the r R's in any of the other $n-1$ positions. Therefore, the probability

that all the red marbles are removed first is

$$
\frac{\binom{n-1}{r}}{\binom{n}{r}} = \frac{\dfrac{(n-1)!}{r!(n-r-1)!}}{\dfrac{n!}{r!(n-r)!}}
$$

$$
= \frac{(n-1)!\,r!(n-r)!}{r!(n-r-1)!\,n!}
$$

$$
= \frac{n-r}{n}
$$

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

Looking ahead to later parts, notice that this probability is equal to the proportion of the marbles that are blue.

(b) Consider the following two events:

- Event X: The final marble that is removed is blue and the final green marble is removed after the final red marble is removed.

- Event Y: The final marble that is removed is green and the final blue marble is removed after the final red marble is removed.

To say that the red marbles were removed first is the same as saying either Event X occurred or Event Y occurred. Since Event X has blue as the final marble and Event Y has green as the final marble, Event X and Event Y cannot occur simultaneously, so this means that the probability the red marbles are removed first is the sum of the probabilities of Event X and Event Y. Therefore, we can compute the desired probability by computing the probabilities of Events X and Y and adding the results together.

For both of these probabilities, we need to count the total number of ways that the marbles can be removed. As in part (a), we will think of an order that the marbles can be removed in as a sequence of r R's, b B's, and g G's. There are $\binom{r+b+g}{r}$ ways to place the r R's. After doing this, there will be $b+g$ empty positions, so there are $\binom{b+g}{b}$ ways to place the b B's. Once the R's and B's are placed, there is no choice of where to place the G's, so the total number of sequences is

$$
\binom{r+b+g}{r}\binom{b+g}{b} = \frac{(r+b+g)!(b+g)!}{r!(b+g)!\,b!g!} = \frac{(r+b+g)!}{r!b!g!}
$$

Note: This expression is an example of something called a *multinomial coefficient*, which is a generalization of a binomial coefficient. We will not use the notation of multinomial coefficients in this solution, but you might like to read about them if you have not before.

We will compute the number of sequences that end in B and have at least one G to the right of the rightmost R. This quantity divided by $\dfrac{(r+b+g)!}{r!b!g!}$ is the probability that Event X occurs.

There are $\binom{r+b+g-1}{b-1}$ ways to place the B's so that the final letter in the sequence is B. This is because we need to place one B in the final position, then place the $b-1$ remaining B's in the remaining $r+b+g-1$ positions.

For any such placement of the B's, in order to place the R's and G's so that at least one G occurs to the right of the rightmost R, we need to place a G in the rightmost empty position. Thus, for every such way to place the B's, we now have $\binom{r+g-1}{g-1}$ ways to place the G's. This is because we must place one G in the rightmost position that is not occupied by a B, and then the remaining $g-1$ G's can be placed in any of the remaining $r+g-1$ positions. Once the B's and G's are placed, there is no choice of where to place the R's, so the probability that Event X occurs is

$$\frac{r!b!g!}{(r+b+g)!}\binom{r+b+g-1}{b-1}\binom{r+g-1}{g-1} = \frac{r!b!g!}{(r+b+g)!} \times \frac{(r+b+g-1)!(r+g-1)!}{(b-1)!(r+g)!(g-1)!r!}$$
$$= \frac{bg}{(r+b+g)(r+g)}$$

By a very similar calculation, it can be shown that the probability that Event Y occurs is

$$\frac{bg}{(r+b+g)(r+b)}$$

and so the probability that all the red marbles are removed from the bag first is

$$\frac{bg}{(r+b+g)(r+g)} + \frac{bg}{(r+b+g)(r+b)} = \frac{bg}{r+b+g}\left(\frac{1}{r+g} + \frac{1}{r+b}\right)$$

(c) In part (b), we showed that $p(r) = \dfrac{bg}{r+b+g}\left(\dfrac{1}{r+g} + \dfrac{1}{r+b}\right)$. Following nearly identical reasoning, it can be shown that

$$p(b) = \frac{rg}{r+b+g}\left(\frac{1}{r+b} + \frac{1}{b+g}\right)$$
$$p(g) = \frac{rb}{r+b+g}\left(\frac{1}{r+g} + \frac{1}{b+g}\right)$$

Using the equations above, we will show that $p(b) - p(r)$ is negative. This result will show that $p(r)$ is greater than $p(b)$. Remember that we are assuming $r < b < g$ in this part.

$$p(b) - p(r) = \frac{rg}{r+b+g}\left(\frac{1}{r+b} + \frac{1}{b+g}\right) - \frac{bg}{r+b+g}\left(\frac{1}{r+g} + \frac{1}{r+b}\right)$$
$$= \frac{g}{r+b+g}\left(\frac{r}{r+b} + \frac{r}{b+g} - \frac{b}{r+g} - \frac{b}{r+b}\right)$$

Since r, b, and g are all positive, the quantity $\dfrac{g}{r+b+g}$ is positive, which means $p(b) - p(r)$ is negative if and only if $\dfrac{r}{r+b} + \dfrac{r}{b+g} - \dfrac{b}{r+g} - \dfrac{b}{r+b}$ is negative. Using a common denominator of $(r+b)(b+g)(r+g)$, we get

$$\frac{r(b+g)(r+g) + r(r+b)(r+g) - b(r+b)(b+g) - b(b+g)(r+g)}{(r+b)(b+g)(r+g)}$$

3

Noting that the denominator $(r+b)(b+g)(r+g)$ is also positive, we have that $p(b) - p(r)$ is negative if and only if the numerator of the fraction above is negative. Manipulating the numerator, we have

$$
\begin{aligned}
&r(b+g)(r+g) + r(r+b)(r+g) - b(r+b)(b+g) - b(b+g)(r+g) \\
&= r(r+g)(b+g+r+b) - b(b+g)(r+b+r+g) \\
&= (r^2 + rg)(r + 2b + g) - (b^2 + bg)(2r + b + g) \\
&= r^3 + 2r^2b + r^2g + r^2g + 2rbg + rg^2 - 2rb^2 - b^3 - b^2g - 2rbg - b^2g - bg^2 \\
&= r^3 + 2r^2b + 2r^2g + rg^2 - 2rb^2 - b^3 - 2b^2g - bg^2
\end{aligned}
$$

At first, it might seem that there is no hope of understanding the expression above. However, if we imagine the situation where $r = b$, then by symmetry we really should have that $p(r) = p(b)$ (in the problem, we are assuming that $r < b$, but the equation above is not aware that we plan to impose this restriction). This means the expression $p(b) - p(r)$ should evaluate to 0 when $r = b$, so it should have a factor of $r - b$. Indeed, if we group the terms that "look" alike, it is easier to notice the factor of $(r - b)$:

$$
\begin{aligned}
&r^3 + 2r^2b + 2r^2g + rg^2 - 2rb^2 - b^3 - 2b^2g - bg^2 \\
&= (r^3 - b^3) + 2(r^2b - rb^2) + 2(r^2g - b^2g) + (rg^2 - bg^2) \\
&= (r-b)(r^2 + rb + b^2) + (r-b)2rb + (r-b)2g(r+b) + (r-b)g^2 \\
&= (r-b)\left[(r^2 + rb + b^2) + 2rb + 2g(r+b) + g^2\right]
\end{aligned}
$$

Since $r < b$ is given, we now have that $p(b) - p(r)$ is negative if and only if the quantity $(r^2 + rb + b^2) + 2rb + 2g(r+b) + g^2$ is positive, but each of r, b, and g is positive, so it is indeed positive!

Therefore, we have that $p(b) - p(r) < 0$, or $p(b) < p(r)$. Similar calculations can be used to show that $p(g) < p(b)$. Therefore, the marble colour with the smallest number of representatives is most likely to be removed first, and the marble colour with the largest number of representatives is least likely to be removed first.

(d) Let $f(r) = \dfrac{r}{r+b+g}$, $f(b) = \dfrac{b}{r+b+g}$, and $f(g) = \dfrac{g}{r+b+g}$. That is, $f(r)$ is the proportion of the marbles that are red, and similarly for $f(b)$ and $f(g)$.

From part (b), the probability that all the red marbles are removed first, $p(r)$, is

$$
\frac{bg}{(r+b+g)(r+g)} + \frac{bg}{(r+b+g)(r+b)}
$$

Dividing the numerator and denominator of each expression by $(r+b+g)^2$, the probability above is equal to

$$
\frac{\dfrac{bg}{(r+b+g)^2}}{\dfrac{r+g}{r+b+g}} + \frac{\dfrac{bg}{(r+b+g)^2}}{\dfrac{r+b}{r+b+g}} = \frac{\dfrac{b}{r+b+g} \times \dfrac{g}{r+b+g}}{\dfrac{r}{r+b+g} + \dfrac{g}{r+b+g}} + \frac{\dfrac{b}{r+b+g} \times \dfrac{g}{r+b+g}}{\dfrac{r}{r+b+g} + \dfrac{b}{r+b+g}}
$$

$$
= \frac{f(b)f(g)}{f(r) + f(g)} + \frac{f(b)f(g)}{f(r) + f(b)}
$$

So we have written the probability that all the red marbles are removed first in terms of the proportions of each of the three colours.

We will now try to explain why the formula above makes sense. Once again, we will think of a way of drawing the marbles randomly as a sequence of R's, B's, and G's. In any given position, the proportion of sequences with a B in that position is $f(b)$. In particular, $f(b)$ is the probability that the final letter in the sequence is B.

The quantity $\dfrac{f(g)}{f(r) + f(g)}$ is equal to the proportion of green marbles among those marbles that are either red or green. Similar to the reasoning above, this means that if we look at sequences of R's, B's, and G's the quantity $\dfrac{f(g)}{f(r) + f(g)}$ is the probability that G is the rightmost letter among R's and G's.

The overall rightmost letter being equal to B is independent of the probability that the rightmost letter among R's and G's is G, so the quantity

$$\frac{f(b)f(g)}{f(r) + f(g)} = f(b) \times \frac{f(g)}{f(r) + f(g)}$$

is equal to the probability that the rightmost letter in the sequence is a B and among R's and G's, the rightmost letter is G. Similarly, the quantity $\dfrac{f(b)f(g)}{f(r) + f(b)}$ is equal to the probability that the rightmost letter is G and among R's and B's, the rightmost letter is B. The R's all occur before the final G and before the final B if and only if one of these two events occurs, so this explains why the probability that all the red marbles are removed first is equal to

$$p(r) = \frac{f(b)f(g)}{f(r) + f(g)} + \frac{f(b)f(g)}{f(r) + f(b)}$$

This expression is in terms of the proportions of red, blue, and green marbles. Similar expressions can be computed for $p(b)$ and $p(g)$.