TA Scheduling for the UW: Introductory Programming Lab

Kaitlyn Arlt, Elizabeth McKinnie, Jierui (Jaron) Wang

30 November 2018

Abstract

The course coordinators of CSE 14x wish to distribute the hours that TAs work in the Introductory Programming Lab (IPL) such that more TAs are working when more students are in attendance, maximizing the number of students helped. The problem was modeled as an apportionment problem, where the hours that TAs work were distributed over populations of how many students are expected to be in attendance for each hour that the IPL is open. Many different methods of the apportionment method were utilized, and we ultimately believe that using the Jefferson method will produce the most applicable distribution of hours.

Project Description

The Introductory Programming Lab (IPL) is a space where students enrolled in the University of Washington's CSE 142 and 143 classes can come to ask questions about their homework. The IPL is staffed by CSE 142 and CSE 143 TAs. When the IPL is open, there are a variety of staffing requirements for operation: a minimum of 1 TA on duty during every hour, and the TAs must work at least one hour at a time, a minimum of two hours every week. Given a fixed number of TA working hours each week during the quarter, our goal is to find a fair distribution of the hours across all hours the IPL is open in order to best meet student demand (how many students come into the IPL and ask questions) and provide high quality tutoring. The problem fits very nicely into an apportionment problem with no simplifications needed.

The Apportionment Problem

Proportional division of a whole-number using fractional ratios is a widely explored topic in mathematics. It would be fortunate if the ratios happened to be exactly integer solutions; however, this is often not the case. An everyday example of such a problem would be the following: A mother has 50 candies, and she wants to give them to her 5 children. An obvious solution would be to give each child 10 candies; however, she wants to get her children to do chores around the house. Instead, she gives each child a number of candies based on the amount of chores they have done. Since she only has 50 candies, she must give a proportion of those 50 based on the proportion of each child's chores to the total chores completed. However, this proportion is likely not a whole-number and the number of candies she gives must be. Whose candy quotas (expected amount per child) does she round down or up? Such a type of problem is generally known as an apportionment problem (AP), a method of solving it is called an apportionment method [1].

A particular context of AP, known as the constitutional apportionment problem (CAP), has been well-established from the constitution of the United States of America. In particular, the fourteenth amendment states that "Representatives shall be apportioned... according to their respective numbers..." and "each state shall have at least one representative" [1]. Thus, the problem is formulated as determining the number of parliamentary representatives a_i for each state i based on population, with at least one representative for each state.

In addition, the distribution must also meet what is called the *quota*. If a state i has population P_i with a country population of P, a house has total members h. With $D = \frac{P}{h}$, then the *exact quota* q_i can be given by $q_i = \frac{P_i h}{P}$ or $\frac{P_i}{D}$. In the rare case where q_i is a whole number, then the obvious solution is $a_i = p_i$. Otherwise, the *lower quota* is q_i rounded down conventionally, formally $\lfloor q_i \rfloor$, and the *upper quota* is q_i rounded up conventionally, formally $\lceil q_i \rceil$. A solution a_i is said to *satisfy lower quota* if $a_i \geq \lfloor q_i \rfloor$ and *satisfy upper quota* if $a_i \leq \lceil q_i \rceil$. A solution is said to *satisfy quota* if it satisfies both upper and lower quota [2]. There are many solutions to the CAP that satisfy quota, so methods were developed in order to find "optimal" solutions based on various metrics.

The Vinton Method

One of the first methods that was proposed to solve the problem was the Vinton method (also referred to as the Hamilton method) [1, 2]. Vinton's method utilizes what is known as the *residue*. The residue of a number is simply the decimal after the number, i.e the residue of 3.7421 is 0.7421. With that in mind, the steps to the Vinton method are as follows: first, calculate the lower quota, $\lfloor p_i \rfloor$, of each state, and assign that many representatives to said state. Then, assign the remaining representatives to states in order of largest to smallest residue [2].

This method is fairly intuitive. The principle behind it is that we want to round up the states which are closest to getting another representative. However, this leads to unintended consequences, particularly the Alabama and new state paradoxes [1]. The Alabama paradox was a particular case where the number of total house representatives went up, but Alabama's number of representatives went down, even though its population remained the same. Similarly, strange behavior occurred when a new state was added (dubbed the new state paradox). Thus, other methods were developed so that such paradoxes could be avoided; these methods are said to be *house-monotone* [2].

Jefferson's Method

One of the proposed issues with Vinton's method was that the assigning of surplus representatives leads to the preferential treatment of some states over others. To counteract this issue, Jefferson proposed a method which removed the need of surplus representatives.

Instead of using the exact quota, Jefferson's method utilizes a *modified quota* which is found by decreasing D such that when you round all the modified quotas down, you get the correct total number of house representatives. More formally, this is finding some d such that the sum of all $a_i = \lfloor \frac{p_i}{D-d} \rfloor$ such that $\sum a_i = h$ [3].

Adams' Method

In response to Jefferson's Method, John Adams proposed his method which was very similar in function, but functionally inverse. Instead of decreasing D and rounding down, Adam's method increases D and rounds up, with the same condition that when the number of representatives of each state is added using this modified quota, the total adds up to h. Formally, we find $a_i = \lceil \frac{p_i}{D+d} \rceil$ such that $\sum a_i = h$ [3].

Both Jefferson's and Adams' methods were turned down for usage in the actual CAP, however, as they were shown to systematically favor either the larger or the smaller states. The reason for this is quite simple: Jefferson's method favors larger states because it increases all the values and rounds down. Thus, the larger values are the one which are increased first. Similarly, with Adams' method, the smaller states get preferential treatment [3].

Dean's Method

When Adams proposed his method as a counter to Jefferson's, James Dean was proposing his method to Daniel Webster for the consideration of Congress [3]. Unlike previous methods, Dean proposed a more mathematical reasoning of determining the rounding criteria for apportionment. His methods are as follows: First, calculate the lower and upper quota. Then find the *harmonic mean* of the two values, given that the harmonic mean of a, b is $\frac{2ab}{a+b}$. If the exact quota exceeds the harmonic mean, then round the number up. Otherwise, round down. Adjust D as necessary so that the total number of representatives apportioned is equal to h.

While not as clearly biased as Adams' method, Dean's method still favors the smaller states. As n gets larger and larger, the harmonic mean of $n, n+1$ approaches $n + 0.5$. For small n, the harmonic mean is less than $n + 0.5$, which results in smaller states having a greater chance of rounding up.

Webster's Method

Webster, after receiving Dean's proposal, realized the evident bias in the method due to the rounding criteria. Thus, he proposed the use of traditional rounding in his method. To apply Webster's, the first step is to calculate the exact quotient [3]. Then, each decimal is rounded in a traditional sense, that is the rounding criteria is the arithmetic mean of the lower and upper quotas. Like the previous methods, this method adjusts D until the sum of all rounded quotas equals h.

The Huntington-Hill Method

Many years after the development of the previous methods, Hill and Huntington proposed a new approach. Their idea was that for any pair-wise comparison of states, no apportionment could be improved by transferring one representative between the pair [2]. Similarly, they decided to compare the previously known methods using pair-wise comparisons and found how in what sense the previous methods were optimal.

They found that the Jefferson method had the minimal value for $|a_i \frac{P_j}{P_i} - a_j|$, where i, j are different states. Likewise, the Adam method optimizes $|a_i - \frac{P_i}{P_j} a_j|$ [4], Dean's method optimizes $|\frac{P_j}{a_j} - \frac{P_i}{a_i}|$ [4], and Webster's method optimizes $|\frac{a_i}{P_i} - \frac{a_j}{P_j}|$ [4].

After analyzing the previous methods, the metric which Huntington decided to utilize for their method was that, for each state, the number of people per representative was as close to that of any other state as possible. Formally, their method minimizes $|\frac{\frac{P_i}{a_i}}{\frac{P_j}{a_j}} - 1|$ [4]. The way they accomplish this is by utilizing the *geometric mean*. Recall that the geometric mean of a, b is $\sqrt{ab}$. The steps behind their method are as follows: first, calculate the exact quota. Then calculate the geometric mean of the lower and upper quotas. Use this number for the rounding criteria. Adjust D as necessary to ensure that the sum of all representatives is h.

Results

The schedule produced by each method is given in the Figures 3 through 6 below. For each figure, the method used is given in the header of the table. The value in each index of the schedule represents the total number of TAs that the particular method assigned to that index. The IPL opens every day at 12:30, and closes at 10:30 on Monday, at 8:30pm on Tuesday, 9:30pm on Wednesday and Thursday, 6:30pm on Friday, and 2:30pm on Saturday and Sunday. To show this, all indexes that are on or after closing times are left blank. The various methods were run with the student distribution data given in Figure 1, such that the value at any index in the distribution table represents the total number of students that came in at that time. Since this data is not currently available to us due to privacy concerns (but is being collected by the CSE department), we created an example data set.The code we used to generate our data and to run the methods can be found at https://github.com/TheNightly/Math381-Project.

Student Distribution							
	Mon	Tues	Wed	Thurs	Fri	Sat	Sun
12:30	319	327	357	369	364	284	265
1:30	375	436	315	455	321	265	247
2:30	348	460	370	442	397		
3:30	393	474	430	420	361		
4:30	407	500	440	498	312		
5:30	443	625	471	565	221		
6:30	418	588	519	566			
7:30	327	349	494	442			
8:30	326		385	357			
9:30	367						

Figure 1: Student Distributions

Exact Quota							
	Mon	Tues	Wed	Thurs	Fri	Sat	Sun
12:30	1.84	1.88	2.05	2.12	2.09	1.63	1.52
1:30	2.16	2.51	1.81	2.62	1.85	1.52	1.42
2:30	2	2.65	2.13	2.54	2.28		
3:30	2.26	2.73	2.47	2.42	2.08		
4:30	2.34	2.88	2.53	2.86	1.79		
5:30	2.55	3.6	2.71	3.25	1.27		
6:30	2.4	3.38	2.99	3.26			
7:30	1.88	2.01	2.84	2.54			
8:30	1.88		2.21	2.05			
9:30	2.11						

Figure 2: Exact Quotas

Hamilton's Method							
	Mon	Tues	Wed	Thurs	Fri	Sat	Sun
12:30	3	3	3	3	3	3	3
1:30	3	4	3	4	3	3	2
2:30	3	4	3	4	3		
3:30	3	4	3	3	3		
4:30	3	4	4	4	3		
5:30	4	5	4	4	2		
6:30	3	5	4	4			
7:30	3	3	4	4			
8:30	3		3	3			
9:30	3						

Figure 3: Hamilton's Method

After reviewing all of these apportionment methods, we decided to consider which metric would be the most suitable for the apportionment of TAs. There are a few main differences

Jefferson's Method							
	Mon	Tues	Wed	Thurs	Fri	Sat	Sun
12:30	3	3	3	3	3	2	2
1:30	3	4	3	4	3	2	2
2:30	3	4	3	4	3		
3:30	3	4	3	3	3		
4:30	3	4	4	4	3		
5:30	4	5	4	4	2		
6:30	3	5	4	4			
7:30	3	3	4	4			
8:30	3		3	3			
9:30	3						

Figure 4: Jefferson's Method

Webster's Method							
	Mon	Tues	Wed	Thurs	Fri	Sat	Sun
12:30	3	3	3	3	3	3	3
1:30	3	3	3	4	3	3	2
2:30	3	4	3	4	3		
3:30	3	4	3	3	3		
4:30	3	4	3	4	3		
5:30	4	5	4	4	2		
6:30	3	4	4	4			
7:30	3	3	4	4			
8:30	3		3	3			
9:30	3						

Figure 5: Webster's Method

Huntington's Method using Geometric Mean							
	Mon	Tues	Wed	Thurs	Fri	Sat	Sun
12:30	3	3	3	3	3	3	3
1:30	3	4	3	4	3	3	2
2:30	3	4	3	4	3		
3:30	3	4	3	3	3		
4:30	3	4	4	4	3		
5:30	4	5	4	4	2		
6:30	3	4	4	4			
7:30	3	3	4	4			
8:30	3		3	3			
9:30	3						

Figure 6: Huntington-Hill Method

to consider. Most importantly, for the CAP, the definition of "fairness" used was "a roughly equal amount of something", whether that be the ratio of population to representative or the difference between those values for different states. However, this type of equality is not necessarily a property we want to impose on an apportion of TAs by student population.

On the contrary, we consider the times when the IPL is busy and when it is calm. On any day, we assume the time it takes to get emergency help from other TAs is relatively fixed. Thus, on a busy day, during this fixed amount of time when a potentially needed TA is missing, more students are not being helped than would be on a calm day when that TA is missing. Thus, we consider it better to have a lower ratio of students to TAs on a busy day; that is, we want the larger populations to be favored in the apparition. The method which accomplishes this the best is Jefferson's method. Thus, we strongly suggest the use of Jefferson's method for the apportionment of TAs by student population (Figure 4).

Improvements and Future Work

Although we came up with a reasonable distribution of working TA hours, it's true that all of the methods we explored are valid and can be considered "fair". To get the most "fair" distribution, we could combine the methods together. For example, we could apply the Jefferson method and the Adams method and then for each hour that the IPL is open, choose the maximum number of TAs from the distributions produced by the Jefferson method and from the Adams method. Although this may produce more slots than we have TAs, since TAs are allowed to work more than 2 hours a week, we could at least suggest that some shifts have more TAs working them.

One method we did not explore is the Balinski Method, which was proposed by Balinski in 1974 to replace the way that Congress apportions its Congressional representatives (and is what Congress uses today to do so.) This method is considered more objectively "fair" than the Huntington-Hill method, and does not introduce the Alabama paradox. We could also examine this method to see how it compares to the other six methods we've applied to our problem.

In the future, we could apply the apportionment problem to fairly distribute the hours of veteran TAs and new TAs, and to fairly distribute the hours of 142 TAs and 143 TAs. A new

TA is a TA who was hired at the beginning of the quarter, and a veteran TA is a TA who has worked as a CSE 142 or CSE 143 TA for at least one quarter prior to the current quarter. These problems are similar to the apportionment problem. For example, we could apportion the veteran TAs, and then assign the new TAs around them according to the previous distribution of TA hours we produced. (Since in general we prioritize veteran TAs, as they have more experience than new TAs, this should produce a more effective distribution than if we were to prioritize new TAs.) Alternatively, we could separately apportion the veteran TAs and the new TAs, and then add the distributions together to decide how many total TAs to assign to a given hour. This would optimize both veteran TAs and new TAs, but may not necessarily optimize the total number of TAs assigned for a given hour. We could apply the same idea to apportioning CSE 142 and CSE 143 TAs, although there isn't a natural prioritization between the two groups.

Conclusions

We were able to suggest a fair distribution of TA hours across the hours that the IPL is open. By modeling our problem as an apportionment problem and applying different methods, we compared the results and found the Jefferson method to be best suited to our particular problem, because Jefferson favors hours with more students, and for the IPL, it's better to have more TAs during hours with more students.

We had to research the apportionment problem and learn about the different methods, and the benefits and drawbacks of each. We were able to come up with a fair distribution using pseudo-randomly generated data, but we know that our conclusions would have been better justified if we knew how many students were helped per hour during other quarters. Although our process itself might not have changed, the actual data may have presented some unique behavior which we did not represent in our generated data.

Acknowledgements

We would like to thank our community partners Cherie Ruan (cruan@uw.edu) and Patty Popp (ppopp20@uw.edu), the CSE 14x course coordinators, for providing us with a great opportunity to solve an interesting and important problem in the real world. We'd also like to thank Professor Sara Billey for helping us rigorously define our problem and suggesting apportionment as a way to solve our problem.

References

[1] Joseph Malkevitch. Apportionment.
`http://www.ams.org/publicoutreach/feature-column/fcarc-apportion1`

[2] M. L. Balinski and H. P. Young. *A New Method for Congressional Apportionment*. Proceedings of the National Academy of Sciences of the USA, 71(11):4602-4606, 1974.

[3] Michael J. Caulfield. *Apportioning Representatives in the United States Congress* Convergence, 2010. DOI:10.4169/loci003163

[4] Joseph Malkevitch. Apportionment II.
`http://www.ams.org/publicoutreach/feature-column/fcarc-apportionii1`

[5] United States Sensus Bureau. Methods of Apportionment.
`https://www.census.gov/history/www/reference/apportionment/methods_of_apportionment.html`

[6] Svante Janson and Svante Linusson. *The Probability of the Alabama Paradox*. Cornell University Library, 2011. `arXiv:1104.2137v2`