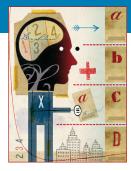# MathWorks Math Modeling Challenge 2021

## Richard Montgomery High School

Team # 14928 Rockville, Maryland

Coach: Warren Wilkerson

Students: Alexander Karbowski, Jerry Shen, Danesh Sivakumar, Tony Wu, Junzhi Xie

## M3 Challenge Technical Computing Award
## RUNNER UP—$2,000 Team Prize

### JUDGES' COMMENTS

*Specifically for Team # 14928 —Submitted at the close of Technical Computing <u>Contention</u> Judging*:

**COMMENT 1:** This paper exemplified quality-over-quantity in its use of technical computing. Python was used to implement a nice "resampling" method to assess model sensitivity in Part 1, and it was used to implement Monte Carlo methods for estimating bandwidth demands in Parts 2 and Part 3. The code was extremely well commented and structured—all models could be easily understood without even referring back to the paper. The students used well-chosen variable and function names, and the judges were happy to see effective code reuse. Notably, the family-level model developed for Part 2 was implemented as a function which was used inside the community-level model developed for Part 3.

*Specifically for Team # 14928 —Submitted at the close of Technical Computing <u>Triage</u> Judging*:

**COMMENT 1:** Reasonable analysis, meaningful use of simulation, and coherent presentation.

**COMMENT 2:** The report is well-written. Nice job in considering both rural and urban communities in US and UK.

**COMMENT 3:** Executive summary is two pages, and it would be better to limit it to one page. Impressive works on justifications and model construction with details on derivation.

# Defeating the Digital Divide: Internet Costs, Needs, and Optimal Planning

## 1 Executive Summary

As the internet continues to garner record amounts of investment and innovation, the world is set to enter a digital revolution, the likes of which have never been seen before. Moreover, the coronavirus pandemic seems to have firmly ingrained our society onto digital platforms, fostering the rapid of growth of internet infrastructure and online opportunity [1]. However, in the wake of this success, many have been left behind. Even in the most developed countries, like the U.K. and U.S., many people with low income and people who live in rural areas find difficulties in accessing the web [5]. Fortunately, the internet's current growth trajectory leaves hope for significant change. In this paper, we will investigate the state of internet accessibility, both in the status quo and in the future, and explore how this transformative service can reach those in need.

As the consideration of internet affordability and equity increasingly becomes significant on a global scale, a mathematical model that predicts the cost of internet for consumers in multiple different environments have never been as relevant and valued. The first model we created predicts the rate of increase in the price of Mbps per dollar and per pound, which allowed us to determine the price of data in the U.S. and U.K. in the next 10 years, up to 2031. Our model relies on the idea that current internet growth trends are set to continue into the future with minimal disruption in the U.S., and will only benefit from the current rollout of 5G in the U.K.. Using public data on internet service costs and average peak bandwidth use from 2015-2021 from both countries, we used Monte Carlo simulations to generate geometric regression models, which allowed us to predict the cost of one megabit of data over the course of the next ten years. Our results show that, by 2031, the cost per unit bandwidth will be 0.00306 USD per Mbps for rural U.K., 0.00076 USD per Mbps for urban U.K., 0.00938 USD for rural U.S., and 0.00469 USD per Mbps for urban U.S.

For the second problem, we created a model to predict the minimum bandwidth that a given household needs to cover their total internet usage 90% and 99% of the time. It is important that people have full access to the internet even when multiple people are using it. Internet activities were split into 5 categories: video conferences at work, web surfing at work, web surfing for leisure, video streams, and games, all of which each require different bandwidths. To determine the amount of hours that are spent on each activity, we used the age and occupation of each person in the household. After finding the hours spent on each activity, we used a Monte Carlo simulation to simulate every possible combination of internet activity. During each run of the simulation, each person was assigned a random activity based on the proportion of how often they did that activity. This was then used to find the total bandwidth that the household uses at that certain point of time. The bandwidths values

1

that were greater than 90% and 99% of the bandwidths represented the required bandwidths necessary to cover the internet usage 90% and 99% of the time. Finally this bandwidth was multiplied by $\frac{100}{65}$ because approximately 35% is lost due to interference. The final results show that for household 1, 16.9 Mbps are required 90% of the time and 20 Mbps are required 99% of the time. For household 2, 10.8 Mbps are required 90% of the time and 15.4 Mbps are required 99% of the time. Finally, for household 3, 13.8 Mbps are required 90% of the time while 16.9 Mbps are required 99% of the time.

Finally, we created a model that was able to designate the number of cell towers necessary in a region based on given region's demographics and specific internet needs. By taking into account how factors like income level, age distribution, and degree of use impacted a region's demand for data flows, we modeled how much bandwidth a region needs using a Monte Carlo simulation. This simulation incorporated the simulation from part 2. However, this time, it also generated random values for the age and the time spent for each activity based on the age distribution for the region. The age gives the time spent online and also the amount of time that is spent working. The time spent in a conference call versus surfing the web varies between 0 and 40 to account for a range of jobs. The results show that 34 towers were needed in area A, 1124 in area B, and 8 in area C. These all support the notion that both income and land mass area influence the bandwidth needed for cell towers.

# Contents

# 2 Problem 1: The Cost of Connectivity

As digital technology solutions advance, there's a burgeoning need for reliable internet usage—no longer is it a convenient commodity; it is absolutely necessary in all spheres of life. With this increase, it is essential to consider a major factor in internet accessibility: affordability. How will the cost of internet use change as time progresses?

## 2.1 Defining the Problem

- We are asked to create a model to predict the cost per unit of bandwidth (in dollars or pounds per Mbps) over the next 10 years for consumers in the United States and the United Kingdom.

## 2.2 Assumptions and Justification

- The current pattern of internet growth will continue for the next 10 years.

    - **Justification:** The expansion of internet will slow eventually due to a physics bottleneck (akin to Moore's law [11]), but this will not occur in the next 10 years.

- Wired broadband is defined as one of the following: cable optic, fiber, DSL (Digital Subscriber Line), or fixed wireless.

    - **Justification:** The listed technologies are the most common, and they have the most data available regarding their costs and bandwidth.

- There is no possibility of 6G being introduced in the U.S. before 2031; thus it will not be considered.

    - **Justification:** The latest technology of 5G has not been fully implemented yet, and and the infrastructure for 6G has not yet been developed. The estimated year for this implementation is at least 2035 [8], which is outside the scope of our model.

- The majority of internet bandwidth cost originates from wired technologies

    - **Justification:** Free internet is much easier to provide with wireless technologies; in addition, approximately 80% of households use wired internet [12].

- Bandwidth speed in rural areas is approximately one-half of that of urban areas.

    - **Justification:** According to bandwidth data provide by Ofcom [9], the bandwidth support in rural areas in the U.K. is around one-half of their urban counterparts; we will be making the assumption that the same can be stated about the U.S.

4

## 2.3 Defining the Variables

| Symbol | Definition | Units |
|---|---|---|
| $t$ | Years since 2015 | Years |
| $C$ | Average monthly cost of internet service before supplementary fees | USD or £ |
| $B_0$ | Average bandwidth consumed in 2015 | Mbps |
| $r$ | Rate of bandwidth increase per year | N/A |
| $G$ | Ratio of average bandwidth of next generation internet to current generation internet | Mbps |
| $a_t$ | Function that returns the adoption rate of next generation internet at year $t$ | % |
| $P_t$ | Function that returns the cost per unit of bandwidth given year $t$ | USD/Mbps or pounds/Mbps |

## 2.4 Developing the Model

### 2.4.1 U.S.

We are to find the cost per unit of bandwidth in dollars per Mbps—this is equivalent to the ratio of monthly cost of internet to the monthly bandwidth consumption in the United States and the United Kingdom. Therefore, we can produce a model that predicts the future trends of cost per bandwidth by discovering a trend in both the monthly cost of internet and the monthly bandwidth consumption and comparing the ratio of these trends.

In order to determine the past trends in the cost of internet, we examined the monthly charge data of the Federal Communications Commission urban rate broadband survey data [2] for the years 2015 to 2021, and likewise the average monthly cost of broadband data for years 2017 to 2020 according to the Worldwide Broadband Price Comparison Data provided by Cable UK [13]. However, we realized there is no general trend over the years in prices present in both of these sources, and therefore we decided to consider the monthly cost of internet as a constant.

Based on observations of plotting the values from the M3 Challenge data set, we decided to opt for an exponential regression to model the increase in bandwidth usage. The model took the following form:

$$y = A \cdot B^t$$

where $y$ is the predicted bandwidth usage, $A$ and $B$ are constants, and $t$ is the number of years since 2015.

From our consideration of the monthly cost as a constant, we now arrive at our cost per unit of bandwidth model for the U.S.:

$$P_t = \frac{C}{B_0 \cdot (1+r)^t}$$

Based on observed trends, it is reasonable to state that the increase in 5G connectivity within the U.S. will not change significantly due to the relatively recent implementation of the technology [8].

### 2.4.2 U.K.

However, the same is not true for the U.K.—5G technology is improving at a considerable rate; thus the adoption rate of the next year's internet $(a_t)$ as well must be considered. Furthermore, the ratio of average bandwidth of next generation internet to current generation internet $(G)$ must be taken into account.

In order to obtain the value of $G$, the following facts given in the M3 data set were utilized:

- The data are for 4G plans, which typically provide download speeds of about **30Mbps**.

- Prices for even 5G connections are only just starting to emerge. In the U.K., these connections provide roughly **200 Mbps** in bandwidth.

Based on this, we can estimate the ratio of average bandwidth of next generation internet to current generation internet by dividing 200 by 30, resulting in the following:

$$G \approx 6.67$$

Finally, we must obtain $a_t$, the adoption rate as a function. This was accomplished by taking the data for Britain's 5G adoption rate from the following graph [4]:



**Fig. 11.** Regional rollout for 2020–2030 based on infrastructure sharing.

The data were imported into an Excel spreadsheet, and a polynomial regression was performed on the values, producing the following equation:

$$a_t = 0.1626(t - 4)^3 - 4.5058(t - 4)^2 + 42.226(t - 4) - 40.397$$

The additional factors for the U.K. were to be added to the denominator of the original model; this is because the corrections affected the amount of bandwidth in a given year and not the fixed monthly cost. Adding the $a_t$ and $G$ terms into our model yielded the following final cost per unit of bandwidth model for the U.K.:

$$P_t = \frac{C}{0.01G \cdot a_t \cdot B_0(1 + r)^t}$$

The graph of 5G adoption in the U.K. demonstrated an interested trend: although access to 5G started to plateau around 2026, the data showed that by 2028, the trend line had started to follow a linear pattern again, and followed this pattern through 2031. This might be because as efforts slow down to expand internet into the remaining, isolated rural areas, governments double down on finishing internet expansion. To model this trend most accurately, we used a fourth-degree polynomial that was able to accurately predict the function on our domain of the years 2015-2031.

## 2.5   Applying the Model

Our model allows for the prediction of bandwidth costs for both rural and urban consumers in the United States and United Kingdom, by substituting respective values that correspond to each scenario.

First, let us determine the 2031 cost per unit bandwidth for rural consumers in the United Kingdom. Since the average cost of monthly internet over the past 5 years is 35.23 USD [13], $C = 35.32$.

To find the constants of the geometric regression that models the trend in the increase of monthly bandwidth usage from 2015, it is not sufficient to just consider the data presented for years 2015 to 2020 [9]. This is because the data sets have a very high standard deviation; for instance, the mean peak download speed for rural U.K. consumers in 2020 is 46.29 Mbps, while the standard deviation is 58.71. To account for this high margin of random error, we conducted a Monte Carlo simulation of 100000 trials. Using the mean and standard deviation for each data set corresponding to years 2015 to 2020, we randomly generated 6 values for the peak download speed (i.e., bandwidth) according the normal distribution of the data in each year. Then, we used this data to create an exponential regression model. By repeating this process of producing random data and creating an exponential regression model through 100000 trials, we are able to determine 10000 distinct values for the parameters in the exponential regression, i.e., the values of $B_0$ and $1 + r$; by taking the average of these values, we are able to find the best exponential regression model with a very limited margin of random

error. This algorithm can be found in the appendix under "Exponential Regression Monte Carlo." The values obtained for $B_0$ and $1+r$ in rural U.K. are 25.1190 and 1.30363, respectively.

Since $t = 16$ when the year is 2031, the value of $a(t)$ is calculated by substituting the value of $t$ into the function, that is,

$$a_4 = 0.1626(16-4)^3 - 4.5058(16-4)^2 + 42.226(16-4) - 40.397 = 98.453$$

Since the average bandwidth of 4G is around 30 Mbps and the average bandwidth of 5G is around 200 Mbps according to data presented by M3 [7], the value of the ratio of average bandwidth of next generation internet to current generation internet is equal to the ratio of 5G bandwidth to 4G bandwidth, i.e., $G = 6.667$.

We have determined the value of each variable in our model. The cost per unit bandwidth is thus equal to

$$P_16 = \frac{C}{(0.01Ga_{16})B_0(1+r)^{16}} = \frac{35.32}{(0.01 \times 6.667 \times 98.453) \times 25.1190 \times (1.30363)^{16}} = 0.00306$$

Therefore, our model predicts that the cost of internet per unit bandwidth is equal to 0.00306 USD, around 0.30 cents per Mbps.

By following the same procedure for both rural and urban populations in the U.K. and U.S., we determine that the cost per unit of bandwidth is **0.00306 USD per Mbps for rural U.K., 0.00076 USD per Mbps for urban U.K., 0.00938 USD for rural U.S., and 0.00469 USD per Mbps for urban U.S.**

Notice that, since we assumed bandwidth coverage in rural U.S. to be 0.5 times that of urban U.S., the cost per unit bandwidth of rural U.S. is equal to exactly 2 times to that of urban U.S.

## 2.6 Evaluation of the Model

The results of our model show that for any area in the U.S. and U.K., the cost per unit bandwidth will be less than 0.01 USD. This is consistent with our expectations; since bandwidth is purchased in bulk per month, as the demand for bandwidth and next generation internet increases exponentially in the next years, the cost and value per unit bandwidth (Mbps) should decrease drastically. As bandwidth usage increases, the unit of bandwidth will probably change, such as to Gbps, but the use of a current unit to report bandwidth results in a very low value for the cost per bandwidth. A similar phenomenon is also seen in commodities; since commodities are purchased in bulk, the cost per unit commodity tends to be a fraction of a cent.

The cost per unit bandwidth was lower in both rural and urban U.K.; this is consistent with data presented by the FCC [2] and Cable UK [13], which show that the costs of internet is significantly lower in the U.K. compared to the U.S.; the U.K. averages at around 35 USD per month compared to the U.S., which averages around 60 USD per month.

The form of our model, a rational function with an exponential regression as a denominator, is also consistent with our expectations about the price of bandwidth. Since all of the values of our variables are positive for $t > 0$, $P_t \neq 0$. As the denominator increases exponentially, the value of the fraction will approach, but never equal, 0. This aligns with the fact that, although the cost per unit bandwidth may drastically decrease, the cost will never equal 0; in other words, bandwidth will never be provided for free.

### 2.6.1   Sensitivity Analysis

As mentioned above, our model incorporated the Monte Carlo simulation to mitigate random errors from our model; this was especially necessary because our original raw data set consisted of only 6 data points, and the standard deviation was very high relative to the mean.

For instance, consider the model with values of $B_0$ and $1 + r$ that result from using the mean of the data presented by FCC for urban U.S. [2]; they are $B_0 = 96.2$ and $1 + r = 1.19821$. Using this model, the predicted cost per Mbps is 0.0425 USD, a value that is 906% of our original value. This is clearly a large margin of error incostistent with our expectations that cost per unit bandwidth will become fraction of a cent, and therefore the Monte Carlo simulation minimizes the random error associated with our model.

## 2.7   Strengths, Weaknesses, and Extensions

Some strengths of this model is that it accepts multiple parameters, and therefore our model can be adjusted to determine the cost per unit bandwidth for multiple scenarios. Although we only included the urban and rural communities of the U.S. and U.K. in this exploration, by simply changing the data set that we consider, we are able to draw conclusions about the cost of bandwidth for any set of people provided that sufficient data is present.

The lack of sufficient data is a major weakness of our model; since data about bandwidth and the internet have been collected only for the past few years, and therefore we are required to make models with relatively few data points. For a small sample size, the trends that we draw are prone to large magnitudes of possible random error. We used the Monte Carlo simulation to mitigate this random error; however, due to the large possibility of error, we are required to run many trials. Around 10000000 trials would have been preferred, however the online compiler used did not have the computational capability to complete these trials in the allotted time frame.

One extension to this model is incorporating the $a_t$ function to the U.S. model as well,

so that we can more accurately track how the introduction of 5G next generation internet changes the average bandwidth usage, thus changing our results. More data about future U.S. plans for development of 5G is required to make this extension. Another extension we can make to the model is to account for future advancements to the internet, such as the introduction of 6G, into the model, and also to combine the multiple models into one cohesive model that can work for any scenario.

# 3    Problem 2: Bit by Bit

In today's society, different groups of people require differing amounts of bandwidth. Younger individuals tend to use greater quantities of bandwidth compared to older individuals. Similarly, occupational roles also play a role in determining the quantity of bandwidth used. Students will likely use the internet frequently resulting in higher bandwidth requirements, while a construction worker will likely need fewer bandwidth requirements. Furthermore, with the onset of virtual meetings in place of the more traditional get-together, bandwidth usage may also increase. This presents an important question: what is the optimal amount of bandwidth necessary to sustain a certain group of individuals. This question will be examined below.

## 3.1    Defining the Problem

- We are asked to create a model that captures the bandwidth that is necessary to sustain a group of individuals 90% and 99% of the time.

- We are asked to apply our model to identify the bandwidth necessary to sustain a group of three sets of individuals.

## 3.2    Assumptions and Justification

- The time that is considered will be the longest time that a person is using bandwidth.

    - **Justification:** When people are not using the internet, there is no bandwidth usage. The times when people are not using bandwidth, such as when sleeping, should not be considered as the question is only relevant when people are using the internet.

- Any individual will only be using bandwidth to perform one task at a time.

    - **Justification:** Individuals will stay focused on the task they are performing whether that be watching TV or doing work online. It is uncommon to find someone streaming a video while also playing video games and while also scrolling through social media.

- People in a single household will perform activities on the internet independently and will not share a single connection.

  - **Justification:** Most people have different tastes. Even among couples, varied interests will result in independent activities which leads to independent bandwidth usage.

- All video streaming will be at the higher definition (HD) resolution.

  - **Justification:** The main goal in this question is to determine the minimum amount of required bandwidth required to sustain daily activities. HD resolution represents a balance between SD resolution and 4K resolution; most consumers would be satisfied with HD resolutions.

- People who work virtually use the internet the entire time they are working.

  - **Justification:** With the COVID-19 pandemic, most white collar jobs have moved to using online communication and platforming. Additionally, these computer jobs almost always must be connected to the internet via a VPN or some other platform.

- When working, the only internet activities people will engage in are video conferences and web surfing.

  - **Justification:** It is socially expected for individuals to work while performing their job. It is typically unacceptable to stream videos or play video games while working.

- The proportion of time people spend on leisure internet activities (streaming, gaming, surfing the web) is the same as before the COVID-19 pandemic.

  - **Justification:** People's interests will not drastically change over the COVID-19 pandemic and they will largely dedicate a similar proportion of their time to the activities they enjoyed before.

## 3.3  Defining the Variables

| Symbol | Definition | Units |
|---|---|---|
| $b(x)$ | The minimum amount of required bandwidth that can cover the total internet $x\%$ of the time. | Mbps |
| $H_t$ | Total Number of Hours Spent Online | Hours |
| $H_e$ | Total Number of Hours Spent at Education/Work | Hours |
| $H_c$ | Hours Spent on Video Conferences at Work | Hours |
| $H_w$ | Hours Spent Surfing the Web at Work | Hours |
| $H_s$ | Hours Spent Surfing the Web for Leisure | Hours |
| $H_v$ | Hours Spent Streaming Video Content | Hours |
| $H_g$ | Hours Spent Playing Games | Hours |
| $b_c$ | Bandwidth for Video Conferences | Mbps |
| $b_s$ | Bandwidth for Surfing the Web | Mbps |
| $b_v$ | Bandwidth for Streaming Video Content | Mbps |
| $b_g$ | Bandwidth for Playing Games | Mbps |

## 3.4  Developing the Model

### 3.4.1  Hours Model

There are two major factors which contribute to the total amount of bandwidth a household uses: bandwidth used while performing occupational tasks and bandwidth used for entertainment. Both of these factors heavily influence how much bandwidth is required by a household. To quantify these ranges, we created a model to capture how age and occupation influence time spent doing web surfing, playing games, video streaming, and engaging in video conferences. This model is shown below:

$$H_e = H_c + H_w$$
$$H_t = H_e + H_s + H_v + H_g$$

$H_c$ and $H_w$ are both dependent on the occupation of the person and must be input manually. Because the work that is done in an occupation typically does not depend on age, it is not included in the calculation. The total amount of time spent on the internet, $H_t$, varies based on age. $H_s$, $H_v$, and $H_g$ will vary based on the average proportion of time each age group has dedicated towards each activity as determined from the internet Media Consumption data-set provided by the M3 Challenge. The ratios of $H_s : H_v : H_g$ and the total time online are shown for each age range below [7] [10]:

| Age | Hours / Week | Surfing | Video | Gaming |
|---|---|---|---|---|
| 2-11 Years Old | 22.33 | N/A | 83.9% | 16.1% |
| 12-17 Years Old | 36.35 | 50.1% | 36.4% | 13.5% |
| 18-34 Years Old | 50.83 | 62.2% | 29.4% | 8.4% |
| 35-49 Years Old | 51.90 | 71.5% | 24.8% | 3.7% |
| 50-64 Years Old | 44.02 | 78.3% | 20.5% | 1.2% |
| 65+ Years Old | 34.12 | 82.6% | 16.9% | 0.5% |

The column Hours/Week describes the quantity of time spent online in total. The columns describing entertainment variables represents the proportion of the time that is used for each value once the hours spent working are subtracted from the Hours/Week.

### 3.4.2 Bandwidth Interaction Model

Ultimately, the amount of bandwidth that is required to sufficiently sustain a group of individuals depends on how much bandwidth is used when a maximal amount of people are online. This occurs when multiple individuals are using the internet concurrently such as when multiple individuals are streaming video or surfing the web.

To quantify when the most bandwidth is being used at once, we run a Monte Carlo simulation where the internet usage of each person is randomly selected. Each person's activity is randomly selected based on a probability distribution proportional to the time people spend on each activity. We can multiply each activity by the bandwidth that it consumes and add for each person in order to get the total bandwidth used at a certain point in time. Letting $n$ be the total number of people in a household yields

$$B = \sum_{i=1}^{n} b_k H_{k_i}$$

for some randomly chosen $k \in \{c, w, s, v, g\}$ with the respective probability distribution of $\left\{ \dfrac{H_{c_i}}{H_{t_i}}, \dfrac{H_{w_i}}{H_{t_i}}, \dfrac{H_{s_i}}{H_{t_i}}, \dfrac{H_{v_i}}{H_{t_i}}, \dfrac{H_{g_i}}{H_{t_i}} \right\}$. Running this Monte Carlo simulation 1 million times yields a distribution of possible bandwidth usages at any point in time when the internet is being used. We want to find the minimum required bandwidth such that $x\%$ of the internet usage is covered at any time in which internet is being used. Thus, we need to find the bandwidth at which $x\%$ of bandwidths in the distribution are smaller.

$$b(x) = \beta : P\left(B < \beta\right) = x$$

In addition, the WiFi speed that a person actually experiences is 20% - 50% less than the advertised bandwidth due to interference and distance from the router [7]. We can take the average of this to get 35%. Correcting for this, the minimum required bandwidth to cover $x\%$ of the internet usage is $b(x) \cdot \frac{100}{65}$.

## 3.5   Applying the Model

**Household 1:** A couple in their early 30s (one is looking for work and the other is a teacher) with a 3-year-old child.

In this example, we have a total of 3 individuals. Their information is summarized in the table below and each of the variables were calculated using the methods described above:

| **Individual** | Age Range | $H_c$ | $H_w$ | $H_s$ | $H_v$ | $H_g$ |
|---|---|---|---|---|---|---|
| Unemployed Person | 18-34 | 0 | 11 | 24.77 | 11.7 | 3.35 |
| Teacher | 18-34 | 12.5 | 27.5 | 6.74 | 3.18 | 0.91 |
| 3 Child | 2-11 | 0 | 0 | 0 | 18.73 | 3.60 |

For the unemployed person, the time spent in conference calls is logically close to 0 as they do not have any conferences to attend when they are unemployed. To find the hours the unemployed person spends surfing the web to look for a job, an estimate was made given the claim that on average people spend 11 hours a week seeking out new jobs [3]. For the teacher, we chose the combination 35 hours in conference calls and 5 hours doing outside web surfing for work given the estimate that teachers are in calls with students 2.5 hours per day 5 days a week. This is based on our own experience with virtual learning in Maryland. Given that teachers typically work 40 hour work weeks, the remaining 27.5 hours are spent doing preparation for class such as creating worksheets and presentations. For the child, they are too young to attend school and therefore only spend time engaging in entertainment online. Based on this information, after running the Monte Carlo simulation, it is expected that the household will need $\boxed{16.9 \text{ Mbps } 90\% \text{ of the time}}$ and $\boxed{20 \text{ Mbps } 99\% \text{ of the time}}$.

**Household 2:** A retired woman in her 70s who cares for two school-aged grandchildren twice a week.

In this example, we similarly have a total of 3 individuals. Because the age of the children is not specified, it was assumed they are between ages 12-17. Their information is summarized in the table below and each of the variables were calculated using the methods described above:

| **Individual** | Age Range | $H_c$ | $H_w$ | $H_s$ | $H_v$ | $H_g$ |
|---|---|---|---|---|---|---|
| Grandmother | 65+ | 0 | 0 | 28.18 | 5.77 | 0.17 |
| Child 1 | 12-17 | 5 | 1.5 | 1.95 | 1.41 | 0.52 |
| Child 2 | 12-17 | 5 | 1.5 | 1.95 | 1.41 | 0.52 |

The grandmother is retired, meaning that she spends no time in conferences or surfing the web for work explaining why a 0 was assigned for both $H_c$ and $H_w$. For the children, because both are in school for 2 days a week at the grandmother's home, they are in conferences for $2 \times 2.5 = 14$ hours a week. For the hours spent surfing the web, this was determined based on our own experience with homework and how long it takes to do homework given the virtual setting. Based on this information, after running the Monte Carlo simulation, it is expected

that the household will need $\boxed{\text{10.8 Mbps 90\% of the time}}$ and $\boxed{\text{15.4 Mbps 99\% of the time}}$.

**Household 3:** Three former M3 Challenge participants sharing an off-campus apartment while they complete their undergraduate degrees full-time and work part-time.

In this example, we similarly have a total of 3 individuals. Their information is summarized in the table below and each of the variables were calculated using the methods described above:

| Individual | Age Range | $H_c$ | $H_w$ | $H_s$ | $H_v$ | $H_g$ |
|---|---|---|---|---|---|---|
| College Student 1 | 18-34 | 20 | 25 | 3.63 | 1.71 | 0.49 |
| College Student 2 | 18-34 | 20 | 25 | 3.63 | 1.71 | 0.49 |
| College Student 3 | 18-34 | 20 | 25 | 3.63 | 1.71 | 0.49 |

Because these individuals are both full-time students and part-time workers, we decided that they would be in video conferences for about 20 hours a week and would engage in online research and surfing around 25 hours a week, giving a total of 45 hours a week of work. This is reasonable assuming these three students are all highly motivated individuals (which is evidenced by the fact that they participated in the M3 Challenge) Based on this information, after running the Monte Carlo simulation, it is expected that the household will need $\boxed{\text{13.8 Mbps 90\% of the time}}$ and $\boxed{\text{16.9 Mbps 99\% of the time}}$.

## 3.6   Evaluating the Model

Based on our model, we determined that each household needs approximately 10–20 Mbps in varying quantities to ensure 90% and 99% coverage. This value makes sense. This is because in many cases, the internet usage will not overlap. In such instances, the maximum amount of bandwidth required is only the process which requires the greatest amount of bandwidth which in our model is the HD streaming with bandwidth usage of 5 Mbps. However, when the internet usage instances overlap, this leads to a greater bandwidth to maintain coverage and thus the higher bandwidth necessary. While these results may seem low initially, they make sense. Because we assumed the only internet processes involved are streaming, video games, surfing, and video conferencing, processes like large downloads are excluded, which typically require much higher levels of bandwidth.

The lower bandwidth requirement for the basic processes outlined in this model suggest that in most cases, the large internet service packages offering bandwidths up to 1000 Mbps are often unnecessary.

### 3.6.1   Sensitivity Analysis

Since we used a Monte Carlo simulation, the model simulates many possible combination of internet usage by all the people. This accounts for all of the randomness possible and every possible bandwidth usage. This randomness allows us to account for both the times when little bandwidth is used and the times with spikes in bandwidth when multiple people are

doing bandwidth-intensive activities, such as streaming.

For the sensitivity analysis, we can vary the bandwidth of certain activities to see how the output will change. One of the biggest factors that affect how much bandwidth one needs is streaming. There is a large difference between HD video streaming and 4K video streaming: 5 Mbps versus 25 Mbps [7]. Using 4K video streaming for each of the 3 households yields

Household 1:
90% coverage: 78.5 Mbps
99% coverage: 81.5 Mbps

Household 2:
90% coverage: 40.0 Mbps
99% coverage: 76.9 Mbps

Household 3:
90% coverage: 13.8 Mbps
99% coverage: 47.7 Mbps

Thus, varying the bandwidth of the possible internet activities changes the required bandwidth drastically. As technology progresses and videos become higher quality, the amount of bandwidth required will increase.

### 3.6.2   Strengths, Weaknesses, and Extensions

One of the major strengths of this model is the incorporation of multiple variables and activities into the calculation of bandwidth usage. Instead of simply taking the average bandwidth usage at any given point in time, the probability of multiple tasks were used instead. This strengthens the model because it highlights the diversity of activities and incorporates the idea that some activities when paired together may drastically increase the amount of bandwidth needed. For example, if multiple people are streaming, our model accounts for this increased bandwidth necessity. On a similar note, incorporating a Monte Carlo simulation of probabilities gives a strong estimate of the differing levels of bandwidth usage at any point in time. By modeling the amount bandwidth at 1 million points in time, it accounts for potential situations where everyone in the household is maxing out on bandwidth while also including potential situations where only one person is performing the simplest task.

One of the major weaknesses in this model is that our model does not account for potential interpersonal connections between individuals in a household. Even though it was assumed in assumption 2-3 that individuals will not share connections, often times, multiple people will spend time streaming or surfing the web together. Our model does not account for this. In the future, we are considering adding a "friend factor" which may allow for this

type of interaction by reducing the amount of time spent using bandwidth alone. Another weakness in this model is that there are many more factors other than age and occupation which play a role in how long an individual spends on the internet. Factors such as income and geographic location play major roles in determining access to the internet and thus can influence the amount of bandwidth needed by a household. This will also be further examined in Problem 3. Finally, this model is largely dependent on the number of hours that a person spends each week online. This was determined via collected data, but much of this data was collected before COVID-19, so this model could be strengthened by using more up-to-date data.

# 4    Problem 3: Mobilizing Mobile

Cell towers are a crucial piece of infrastructure, necessary to provide bandwidth to users so individuals can easily access the internet while away from home. However, different areas require different quantities of bandwidth as a result of various factors including cost of living, geography, and age group. Given the necessity to develop new cell tower infrastructure to support a growing nation, a model will be developed in this section to examine the optimal number of cell towers that should be placed in a given area based on several criteria.

## 4.1    Defining the Problem

- We are asked to create a model that determines the optimal number of cell towers that should be placed given demographic and population data.

- We are asked to apply our model to three different regions.

## 4.2    Assumptions and Justifications

- All people are evenly distributed in each sub-region.

    - **Justification:** Population densities will be restrained by the fact that people do not want to live too close to one another. Without knowledge of other factors that may cause people to cluster together, like the presence of economic hubs or living centers, it is logical to assume that people would spread out naturally.

- Only 4G cell towers will be placed in each location.

    - **Justification:** 4G cell towers have the most research in terms of the their capabilities in the status quo. 5G towers are still an emerging technology and, as a result, cannot be evaluated as effectively. 4G towers are also currently the most ubiquitous model and thus would be the most likely form of cell tower implementation in these regions.

-     - **Justification:**

- The likelihood of an individual or family in a particular sub-region to purchase internet of a certain quality or even at all is tied to the sub-region's median income. A 1% increase in the price of internet is associated with a 1% decrease in affordability.

    - **Justification:** People with more limited pools of revenue will not be able to purchase high quality internet or even internet at all.

- The amount of cell tower-supplied internet used by kids under the age of 7 is negligible.

    - **Justification:** Most children under the age of 7 do not own cell phones or other devices that can actively connect to cell towers.

- The likelihood of an individual or family in a particular sub-region to purchase internet of a certain quality or even at all is tied to the sub-region's median income.

    - **Justification:** People with more limited pools of revenue will not be able to purchase high quality internet or even internet at all.

- "Low income" is defined as less than 40 thousand dollars per year, "middle income" is defined as 40 to 122 thousand dollars per year, and "high income" is defined as more than 122 thousand dollars per year.

    - **Justification:** These are the standard values used by various analytical firms [6].

## 4.3   Defining the Variables

| Variable | Definition | Units |
|---|---|---|
| $L$ | Total area of sub-region | miles$^2$ |
| $C$ | Cost of Living Index [0,100] | N/A |
| $R_T$ | Range of Cell Tower (10 Miles) | Miles |
| $C$ | Cost of Living Index | N/A |
| $R_C$ | Range of cell tower | miles |
| $B_C$ | Bandwidth that can be provided by a cell tower (50 Mbps) | Mbps |
| $U$ | Internet Usage for a Household Given Age Group Distribution | Mbps |
| $T_n$ | The Total Number of Cellular Towers Needed | N/A |

## 4.4   Developing the Model

There are various factors that can contribute to the amount of bandwidth a region needs and therefore the quantity of cell towers needed to cover such area. Age demographics, cost of living, and income all play a role in determining the amount of bandwidth a household needs and can afford. To quantify how these variables all interact, we developed a model which incorporates all of these factors. This model is shown below:

$$T_n = \left\lfloor \frac{L \cdot \sum U}{\frac{C}{100} \cdot \pi(R)^2 \cdot B_c} \right\rfloor$$

This model essentially represents the amount of bandwidth needed per unit land divided by the coverage from each of the towers. The numerator of the function $L \cdot \sum U$ represents the bandwidth needed by all households in the region multiplied by the land area. This gives a general idea of how much bandwidth is needed in total. The denominator represented by $C \cdot \pi(R)^2 \cdot B_c$ represents the total bandwidth provided by the area coverage of the cell towers. The $C$ element is an added function that also takes into account the fact that not every individual in an area has the financial resources to afford cellular data. When the cost of living index is greater than 100, it suggests that items are $\frac{C}{100}$ times more expensive than average. As a result, based on the assumptions above, this value was included in the denominator.

## 4.5   Applying the Model

To find the value of $U$, it is necessary to identify the major factors that contribute to a household's need for bandwidth. Luckily, this can be identified using the model created in the second problem. In this simulation, age was randomly generated following the probability distribution based on the data from each sub-region. Additionally rather than setting the employment values manually as was the case in the previous mode, these were also generated randomly as it represents a range of occupations which various individuals may have. Finally, average household was assumed to be 3 which was calculated given the average population and number of households. Once these were identified, the bandwidth requirements for 1 million such households were generated using a Monte Carlo simulation. The average bandwidth requirement was then found and multiplied by the number of households to deduce total bandwidth needed. The results are shown below.

| Region | $L$ | $\sum U$ | $C$ | $T_n$ |
|---:|---|---|---|---|
| A | 6.83 | 61243 | 80.2 | 34 |
| B | 33.64 | 468646 | 89.3 | 1124 |
| C | 1.64 | 84150 | 124.5 | 8 |

## 4.6   Evaluating the Model

Based on the simulation, 34 towers are required in region A, 1124 towers are required in region B, and 8 towers are required in region C. It is clear that the most number of cellular towers are needed in Region B, given the low cost of living and the high land mass that needs to be covered. In combination, this allows greater access to cell coverage. The region which required the least number of cell towers was region C. This was also unsurprising given its very small land area and the high cost of living. It was also unsurprising that region C had

the least number of cell towers needed, given that it had the highest cost of living index and the highest population density out of the three. One benefit to this model is that it uses a simulation with a significant number of iterations. This allows for the consideration of multiple occupations while also accurately capturing the age ranges for each area.

### 4.6.1 Strengths, Weaknesses, Extensions

A strength of the model is that it was able to predict which regions needed more cellular towers relative to the other regions based on the factors that were to be considered. For example, land area and cost of living both were included in this model, which accurately represents the real world.

A weakness of the model is its consistent overestimate from the simulation; this is due to our original assumption that the consumers are using the internet all the time, which is clearly not the case. Unlike the intention of the model developed for second question, when using the model for this example, it is unknown how long each individual is working for. Thus, our model likely overestimated the amount of time people spend online and consequently the bandwidth required.

A potential extension of the model would address the time consumers generally spend using the internet; this is certainly an important factor that was not completely considered in this model and was overestimated. Furthermore, age and median household income in subareas was ignored because of the time constraints. These would a good way to further increase the resolution of the model.

# 5  Appendix

## 5.1  Exponential Regression Monte Carlo

```python
import numpy as np
from sklearn.linear_model import LinearRegression

# Number of trials for the Monte Carlo simulation
trials = 100000

# The mean bandwidth per month in the years 2015 to 2020, reverse chronological
    order; data for urban U.K. used
MbpsMeans = [97.44247228,
104.1279867,
90.35112135,
86.89859683,
53.61106073,
32.35380689]

# The standard deviation of bandwidth per month in the years 2015 to 2020,
    reverse chronological order; data for urban U.K. used
MbpsStdDev = [102.492798,
102.8019483,
95.61392489,
84.7818465,
45.03451962,
53.77606575
]

# Stores the randomly generated values for bandwidth per month, corresponding to
    the years 2020 to 2015, respectively
MbpsRandValues = [0, 0, 0, 0, 0, 0]

# Stores the natural log of the values in MbpsRandValues
Lny = [0, 0, 0, 0, 0, 0]

# Compounds the value returned for the base of the exponential regression in
    each trial
SumBase = 0
# Compounds the value returned for the coefficient of the exponential regression
    in each trial
SumCoef = 0

# Values of t corresponding to the years 2020 to 2015, respectively
```

```python
years = [5,4,3,2,1,0]

# Loops Monte-Carlo simulation for number of selected trials
for j in range(0, trials):

  # Assigns randomly generated values according to the mean and standard
     deviation of each year
  for i in range(0,6):
    MbpsRandValues[i] = np.random.normal(MbpsMeans[i], MbpsStdDev[i])
    # The standard deviation allows for values of bandwidth that are less than 0;
       however, since the realistic minimum for bandwidth is 0.5 Mbps, all data
       below this value is replaced with 0.5 Mbps
    if (MbpsRandValues[i]<0.5):
      MbpsRandValues[i] = 0.5
    # We take the natural log of the randomly generated values, to transform the
       exponential regression model of coefficient 'a' and base 'b' to a linear
       regression model of intercept ln('a') and coefficient ln('b')
    Lny[i] = np.log(MbpsRandValues[i])

  # Defining lists to be used for the linear regression
  x = np.array(years).reshape((-1, 1))
  y = np.array(Lny)
  # Exponentiating the values of the intercept (ln('a')) and coefficient
     (ln('b')) returns the values of the coefficient of the exponential
     regression ('a') and the base ('b'), respectively
  SumBase += np.exp(LinearRegression().fit(x,y).coef_)
  SumCoef += np.exp(LinearRegression().fit(x,y).intercept_)

# Divides by the total number of trials to return the arithmetic mean
print(SumBase/trials)
print(SumCoef/trials)
```

## 5.2   Bandwidth Monte Carlo

```python
from random import random

def monte_carlo(people, SIMULATIONS, coverage):
    # Most time spent online
    total = max([sum(i) for i in people])

    # Mpbs of each activity (b_c, b_w, b_s, b_v, b_g)
    MBPS = [3, 1, 1, 5, 2, 0]

    values = []
```

```python
# Adds a column for no Internet usage
for person in people:
    person.append(total - sum(person))

# Monte Carlo simulation for bandwidth used at any point in time
for i in range(SIMULATIONS):
    value = 0

    for person in people:
        # Cumulative probability
        cumulative = []
        for p in person:
            previous = 0 if len(cumulative) == 0 else cumulative[-1]
            cumulative.append(p/total + previous)

            # Randomly choses an activity based on proportion of time spent
            chance = random()
            index = 0
            while chance > cumulative[index]:
                index += 1

            # Adds the bandwidth used to the total
            value += MBPS[index]

    # Bandwidth used in that point of time
    values.append(value)

# Count of how many times each bandwidth occurs
counts = {}
for i in values:
    if i not in counts:
        counts[i] = 1
    else:
        counts[i] += 1

# Count of how many times the bandwidth is less than or equal to each
    bandwidth value
cumulative = []
keys = sorted(counts)
for key in keys:
    previous = 0 if len(cumulative) == 0 else cumulative[-1]
    cumulative.append(counts[key] + previous)

# Calculates when the bandwidth is covered 90% and 99% of the time
```

```python
    mbps = []
    for i in coverage:
        index = 0
        while cumulative[index] < SIMULATIONS * i/100:
            index += 1

        # Multiply by 100/65 because the bandwidth will be approximately 35%
            slower due to interferences
        mbps.append(round(keys[index] * (100/65), 1))

    return mbps


SIMULATIONS = 1000000


household1 = [[0, 11, 24.77, 11.7, 3.35],
             [12.5, 27.5, 6.74, 3.18, 0.91],
             [0, 0, 0, 18.73, 3.60]]


household2 = [[0, 0, 28.18, 5.77, 0.17],
             [5, 1.5, 1.95, 1.41, 0.52],
             [5, 1.5, 1.95, 1.41, 0.52]]


household3 = [[20, 25, 3.63, 1.71, 0.49],
             [20, 25, 3.63, 1.71, 0.49],
             [20, 25, 3.63, 1.71, 0.49]]


coverage = [90, 99]


print("Household 1:")
mbps = monte_carlo(household1, SIMULATIONS, coverage)
for i in range(2):
    print(str(coverage[i]) + "% coverage: ", mbps[i], "Mbps")

print("\nHousehold 2:")
mbps = monte_carlo(household2, SIMULATIONS, coverage)
for i in range(2):
    print(str(coverage[i]) + "% coverage: ", mbps[i], "Mbps")

print("\nHousehold 3:")
mbps = monte_carlo(household3, SIMULATIONS, coverage)
for i in range(2):
    print(str(coverage[i]) + "% coverage: ", mbps[i], "Mbps")
```

## 5.3   Age Bandwidth Monte Carlo

This Monte Carlo simulation uses the function monte_carlo() from the subsection above.

```python
import numpy
from random import uniform

hours = [22.33, 36.35, 50.83, 51.90, 44.02, 34.12]
ratios = [[0, 0.839, 0.161],
          [0.501, 0.364, 0.135],
          [0.622, 0.294, 0.084],
          [0.715, 0.248, 0.037],
          [0.783, 0.205, 0.012],
          [0.826, 0.169, 0.005]]

# Distribution for each age group
region1 = [0.115, 0.141, 0.239, 0.206, 0.114, 0.185]
region2 = [0.1024, 0.1273, 0.1452, 0.2292, 0.1488, 0.2471]
region3 = [0.0852, 0.0861, 0.3339, 0.2713, 0.1032, 0.1203]

SIMULATIONS = 10000

def age_monte_carlo(region, SIMULATIONS):
    total = 0
    for i in range(SIMULATIONS):
        household = []
        for j in range(3):
            age = numpy.random.choice(numpy.arange(0, 6), p=region)
            time = hours[age]

            H_c, H_w = 0, 0
            if age == 1:
                H_c = uniform(0, 20)
                H_w = 20 - H_c
            if age in [2, 3, 4]:
                H_c = uniform(0, 40)
                H_w = 40 - H_c

            free = time - H_c - H_w
            H_s = free * ratios[age][0]
            H_v = free * ratios[age][1]
            H_g = free * ratios[age][2]

            household.append([H_c, H_w, H_s, H_v, H_g])

        total += monte_carlo(household, SIMULATIONS, [100])[0]
```

```python
    print(total / SIMULATIONS)

print("Region 1:")
age_monte_carlo(region1, SIMULATIONS)

print("\nRegion 2:")
age_monte_carlo(region2, SIMULATIONS)

print("\nRegion 3:")
age_monte_carlo(region3, SIMULATIONS)
```

# Works Cited

[1] BBC. *Coronavirus: How the world of work may change forever*, 2021.
https://www.bbc.com/worklife/article/20201023-coronavirus-how-will-the-pandemic-change-the-way-we-work.

[2] Federal Communications Commission. *Urban Rate Survey Data & Resources*, 2021.
https://www.fcc.gov/economics-analytics/industry-analysis-division/urban-rate-survey-data-resources.

[3] Alison Doyle. *How Much Time Should You Spend on a Job Search?*, 2020.
https://www.thebalancecareers.com/how-much-time-to-spend-on-a-job-search-2062204.

[4] Zoraida Frias, Edward J. Oughton. *The cost, coverage and rollout implications of 5G infrastructure in Britain*, 2017.
https://www.sciencedirect.com/science/article/pii/S0308596117302781.

[5] Ben Wieder, Chris Zubak-Skees, Ellen Holmes, Eleanor Fox. *Rich people have access to high-speed interner; many poor people still don't*, 2016.
https://publicintegrity.org/inequality-poverty-opportunity/rich-people-have-access-to-high-speed-internet-many-poor-people-still-dont/

[6] Investopedia. *Which Income Class Are You?*, 2019.
https://www.investopedia.com/financial-edge/0912/which-income-class-are-you.aspx.

[7] M3. *Defeating the digital divide data, Mathworks Math Modeling Challenge* 2021, 2021.
https://m3challenge.siam.org/node/523.

[8] Jane McCallion. *What is 6G and how far are we from rollout?*, 2020.
https://www.itpro.co.uk/infrastructure/network-internet/357153/what-is-6g-and-how-far-are-we-from-rollout.

[9] Office of Communications. *UK home broadband performance, November 2015: The performance of fixed-line broadband delivered to UK residential consumers*, 2016. https://www.ofcom.org.uk/research-and-data/telecoms-research/broadband-research/home-broadband-performance.

[10] Victoria Rideout. *The Common Sense Census: Media Use by Tweens and Teens*, 2019. https://www.commonsensemedia.org/sites/default/files/uploads/research/2019-census-8-to-18-full-report-updated.pdf.

[11] Dean Takahashi. *Copper wires might be the bottleneck in the way of Moore's law*, 2012. https://venturebeat.com/2012/12/11/copper-wires-might-be-the-bottleneck-in-the-way-of-moores-law

[12] Ann Treacy. *Wired vs wireless home connections—20 percent use wireless*, 2017. https://blandinonbroadband.org/2017/03/20/wired-vs-wireless-home-connections-20-percent-use-wireless

[13] Cable UK. *Worldwide Broadband Price Research*, 2020.
https://www.cable.co.uk/broadband/pricing/worldwide-comparison/#resources.