# Parameter Estimations and Their Applications

A report submitted in partial fulfillment for the requirements

of the degree of Bachelor's of Science in Mathematics

Kyle Kibodeaux

Under the guidance of

Jebessa Mijena, Phd.

12/14/2018

Georgia College and State University

Table of Contents

Acknowledgements

I would like to thank Dr. Mijena for being a great capstone advisor and for all his assistance with the material. I would also like to thank the Mathematics Department of Georgia College and State University for their continued support and assistance through my undergraduate career. I would also like to thank my friends and family for all their support.

**Abstract**

The purpose of this research involves a focus on parameter estimations and their asymptotic properties for various discrete and continuous distributions and their real-life applicability. We estimated distributions of the service times to get food, inter-arrival time, and the number of people arriving during lunch hour between 12:00 pm to 1:00 pm at the Bobcat Food Court's Chick-fil-A and Subway and calculated the estimate of their parameters. We then discovered that the service times for Chick-fil-A and Subway could be modeled by a weilbull distribution, while the inter-arrival times for both followed the exponential distribution. Our results show that the average service time for Chick-fil-A is significantly longer than Subway.

# 1 Introduction

Parameters are descriptive measures of an entire population; however, it is usually impossible to get data on the entire population due to limited resources. Therefore, instead of trying to get data from the whole population, we take samples and then estimate the parameters for the population. One estimation technique that is used often is maximum likelihood estimation (MLE). MLE is an effective method because it can estimate the most probable parameter based on sample observations alone. Since MLE can be used from observations alone, we can then conduct our own study and analyze it in RStudio. Our study will be conducted at the Bobcat Food Court, which contains a Chick-fil-A and Subway. For Chick-fil-A and Subway, we will find the service time, inter-arrival time, and the number of people that go to both during the hour; then, we will find the model that fits the data the best and estimate their parameters using MLE. Once this is done, we will compare the service time between Chick-fil-A and Subway to see which one on average takes longer.

# 2 Maximum Likelihood Estimation

In order to understand MLE, we need to first understand what a likelihood function is. Let us consider having a random sample $X_1, X_2 ... X_n$ and an unknown parameter $\theta$. We get that the probability mass function (pmf) of the one-parameter function is given by $f(x;\theta)$. It follows that the joint pmf is given by

$$\begin{aligned} L(\theta) &= f(x_1;\theta)f(x_2;\theta)f(x_3;\theta)...f(x_n;\theta) \\ &= \prod_{i=1}^{n} f(x_i;\theta) \end{aligned} \tag{1}$$

$L(\theta)$, which is the joint pmf, is also called the likelihood function for a one-parameter function. The likelihood function describes, given the data, the likelihood of the estimated parameter. Now that we know what the likelihood function is, we want to find the parameter that maximizes the likelihood function so we can get the best possible estimate of the parameter. Since the likelihood function is the product of multiple functions, it can become very hard to maximize, Therefore to simplify we take the logarithm of the likelihood function which is called the log-likelihood function.

This is allowed since a logarithm is a monotonic increasing function, thus anything that maximizes the likelihood function also maximizes the log-likelihood function. Often we denote the log-likelihood as $\mathcal{L}(\theta)$. By taking the derivative of this function we can now solve for its maximum. Thus, the MLE can be solved by

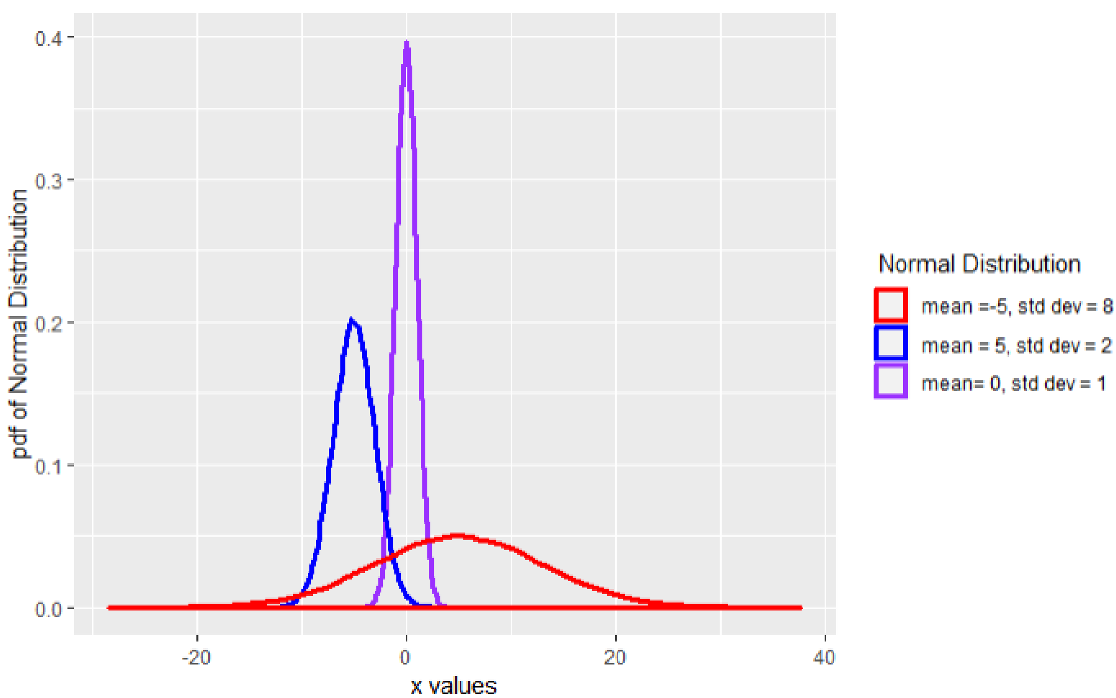
$$\frac{d[\mathcal{L}(\theta)]}{d\theta} = 0 \tag{2}$$

Similarly, the likelihood function with n-parameters is given by

$$\mathcal{L}(\theta_1, \theta_2, \theta_3, ..., \theta_n) = \prod_{i=1}^{n} f(x_i; \theta_1, \theta_2, \theta_3, ..., \theta_n) \tag{3}$$

Now we will look at some distributions and find the MLE for each.

## 2.1 Normal Distribution




The normal distribution is the most common distribution and is often referred to as the bell-shape curve; its pdf is given by

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$$

where $-\infty < x < \infty$ with parameters $\mu$ and $\sigma$ where $-\infty < \mu < \infty$ and $0 < \sigma < \infty$. . The normal distribution has two parameters: the mean, $\mu$ which is the location parameter, and standard deviation, $\theta$ which is the scale parameter. Often we will see a Standard normal distribution which is a normal distribution with $\mu = 0, \sigma = 1$ where the pdf of this is $\frac{1}{\sqrt{2\pi}}e^{\frac{-x^2}{2}}$ Many random variables follow a normal distribution such as height, weight, grade distribution, and IQ.

We will now solve for the MLE of the normal distribution. Note that the likelihood function is given by

$$L(\mu, \sigma) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{[\frac{-(x_i-\mu)^2}{2\sigma^2}]}$$

Thus the log-likelihood is given by

$$\begin{aligned} \mathcal{L}(\mu, \sigma) &= \ln\left(\prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(x_i-\mu)^2}{2\sigma^2}}\right) \\ &= ln \sum (\frac{1}{\sqrt{2\pi\sigma^2}}) + ln(e^{\sum - \frac{(x_i-\mu)^2}{2\sigma^2}}) \\ &= -\frac{n}{2} ln(2\pi\sigma^2) - \sum_{i=1}^{n} \frac{-(x_i-\mu)^2}{2\sigma^2} \\ &= \frac{-n}{2} ln(2\pi) - nln\sigma - \sum_{i=1}^{n} \frac{-(x_i-\mu)^2}{2\sigma^2} \end{aligned} \tag{4}$$

Now to maximize this function we need to take two partial derivatives in respect to both the parameters.

$$\begin{aligned} \frac{\partial}{\partial\mu}(\mathcal{L}(\mu, \sigma)) &= \frac{\partial}{\partial\mu}(\frac{-n}{2} ln(2\pi) - nln\sigma - \sum_{i=1}^{n} \frac{-(x_i-\mu)^2}{2\sigma^2}) \\ &= 0 - 0 + \frac{1}{\sigma^2}(\sum_{i=1}^{n} x_i - n\mu) \end{aligned} \tag{5}$$

Now to find the maximum

$$\begin{aligned} 0 &= \frac{1}{\sigma^2}(\sum_{i=1}^{n} x_i - n\mu) \\ &= \sum_{i=1}^{n} x_i - n\mu \end{aligned} \tag{6}$$

Solving for $\mu$ we get that

$$\mu = \frac{\sum_{i=1}^{n} x_i}{n} = \bar{x} \tag{7}$$

Now take the second derivative with respect to $\mu$ and we get a negative value, thus the MLE of $\mu = \bar{x}$

Now we must take the partial derivative with respect to $\sigma$

$$\begin{aligned} \frac{\partial}{\partial \sigma}(\mathcal{L}(\mu,\sigma)) &= \frac{\partial}{\partial \sigma}\Big(\frac{-n}{2} ln(2\pi) - n ln \sigma - \sum_{i=1}^{n} \frac{-(x_i-\mu)^2}{2\sigma^2}\Big) \\ &= 0 - \frac{n}{\sigma} + \frac{1}{\sigma^3}\Big(\sum_{i=1}^{n}(x_i-\mu)^2\Big) \end{aligned} \tag{8}$$

Solving for the maximum, we get

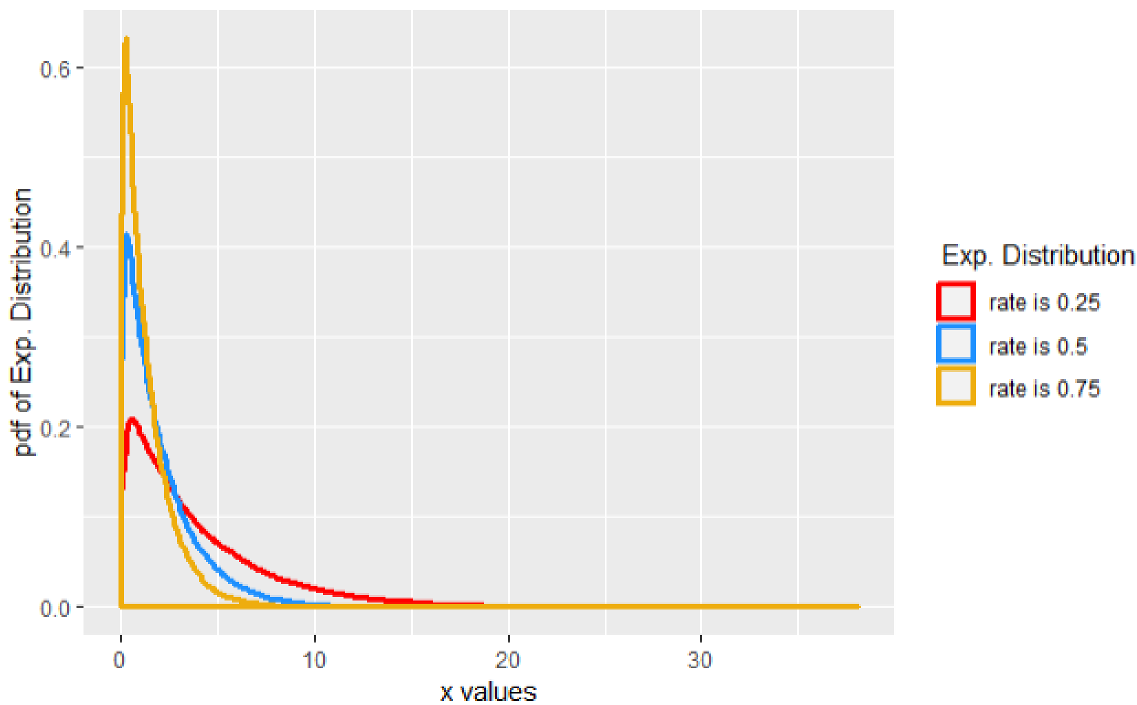
$$\begin{aligned} 0 &= -\frac{n}{\sigma} + \frac{1}{\sigma^3}\Big(\sum_{i=1}^{n}(x_i-\mu)^2\Big) \\ &= -n + \frac{1}{\sigma^2}\Big(\sum_{i=1}^{n}(x_i-\mu)^2\Big) \end{aligned} \tag{9}$$

Solving for $\sigma$, we get that

$$\begin{aligned} \sigma^2 &= \frac{\sum_{i=1}^{n}(x_i-\mu)^2}{n} \\ &= \frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n} \end{aligned} \tag{10}$$

Note that the second derivative $< 0$. Thus the MLE of $\sigma = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n}}$.

## 2.2 Exponential Distribution




The exponential distribution is often used to model the amount of time between events with a rate parameter given by $\frac{1}{\theta}$. It is often used to test the reliability of a product. The pdf of the exponential distribution is given by

$$f(x) = \frac{1}{\theta} e^{\frac{-x}{\theta}}$$

where $0 < x < \infty, \theta \in \Omega = \{\theta : 0 < \theta < \infty\}$

Thus the likelihood function of the exponential distribution is given by

$$\begin{aligned} L(\theta) &= \prod_{i=1}^{n} \frac{1}{\theta} e^{\frac{-x}{\theta}} \\ &= \frac{1}{\theta^n} exp\left(\frac{-\sum_{i=1}^{n} x_i}{\theta}\right) \end{aligned} \tag{11}$$

Taking the logarithm of both sides leaves us with the log-likelihood function

$$\begin{aligned} \mathcal{L}(\theta) &= ln\big(\frac{1}{\theta^n}exp\big(\frac{-\sum_{i=1}^n x_i}{\theta}\big)\big) \\ &= -nln(\theta) - \frac{1}{\theta}\sum_{i=1}^n x_i \end{aligned} \tag{12}$$

Now taking the derivative with respect to theta, we get

$$\begin{aligned} \frac{d}{d\theta} &= \frac{d}{d\theta}\big( - nln(\theta) - \frac{1}{\theta}\sum_{i=1}^n x_i\big) \\ &= \frac{-n}{\theta} + \frac{\sum_{i=1}^n x_i}{\theta^2} \end{aligned} \tag{13}$$

Now to find the maximum we set $= 0$

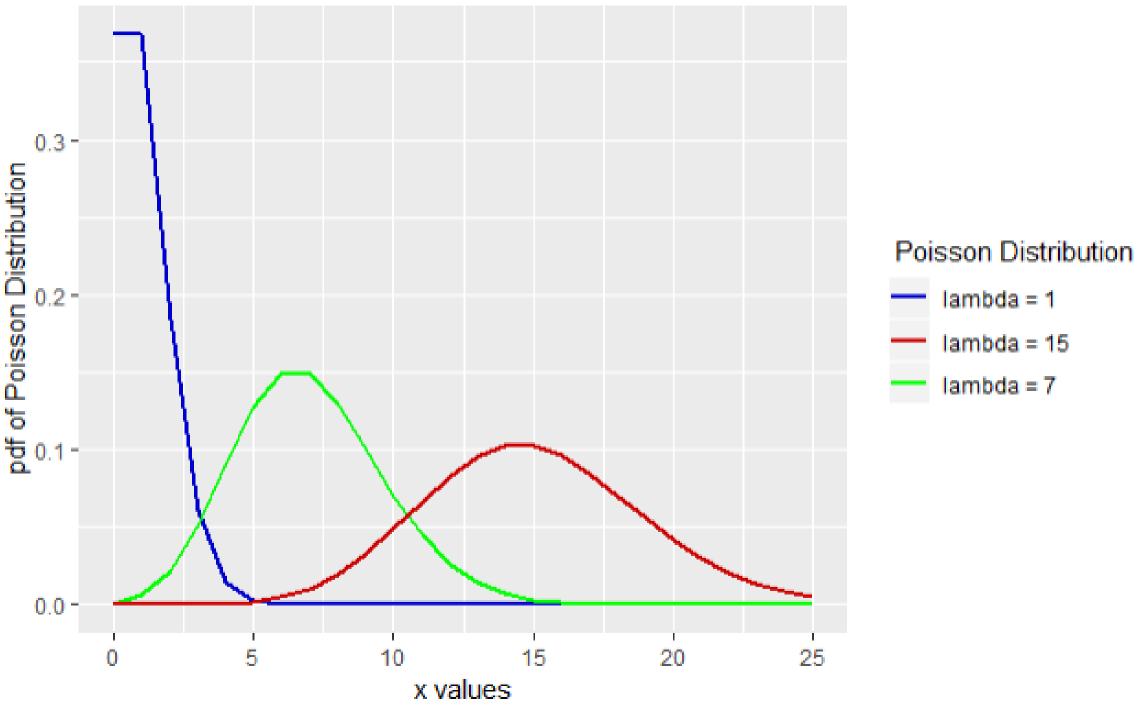
$$\begin{aligned} 0 &= \frac{-n}{\theta} + \frac{\sum_{i=1}^n x_i}{\theta^2} \\ &= -n + \frac{\sum_{i=1}^n x_i}{\theta} \end{aligned} \tag{14}$$

Solve for $\theta$ to obtain

$$\theta = \tfrac{1}{n}\sum_{i=1}^n x_i = \bar{x}$$

Note that the second derivative $< 0$, thus the MLE of $\theta = \bar{x}$

## 2.3 Poisson Distribution




The Poisson distribution expresses the frequency of an event in a given amount of time. It is a one-parameter function where $\lambda$ is the rate parameter. The pdf of the Poisson distribution is given by

$$f(x) = \frac{e^{-\lambda}\lambda^x}{x!}$$

Thus it follows that the likelihood function is given by

$$L(\lambda) = \prod_{i=1}^{n} \frac{e^{-\lambda}\lambda^{x_i}}{x_i!}$$

Thus, the log-likelihood functions is

$$\begin{aligned}\mathcal{L}(\lambda) &= ln(\prod_{i=1}^{n} \frac{e^{-\lambda}\lambda^{x_i}}{x_i!}) \\ &= \sum_{i=1}^{n} ln(\frac{e^{-\lambda}\lambda^{x_i}}{x_i!}) \\ &= \sum_{i=1}^{n}(-\lambda + x_i ln(\lambda) - ln(x_i!)) \\ &= -n\lambda + \sum_{i=1}^{n} x_i ln(\lambda)\end{aligned} \tag{15}$$

It follows that

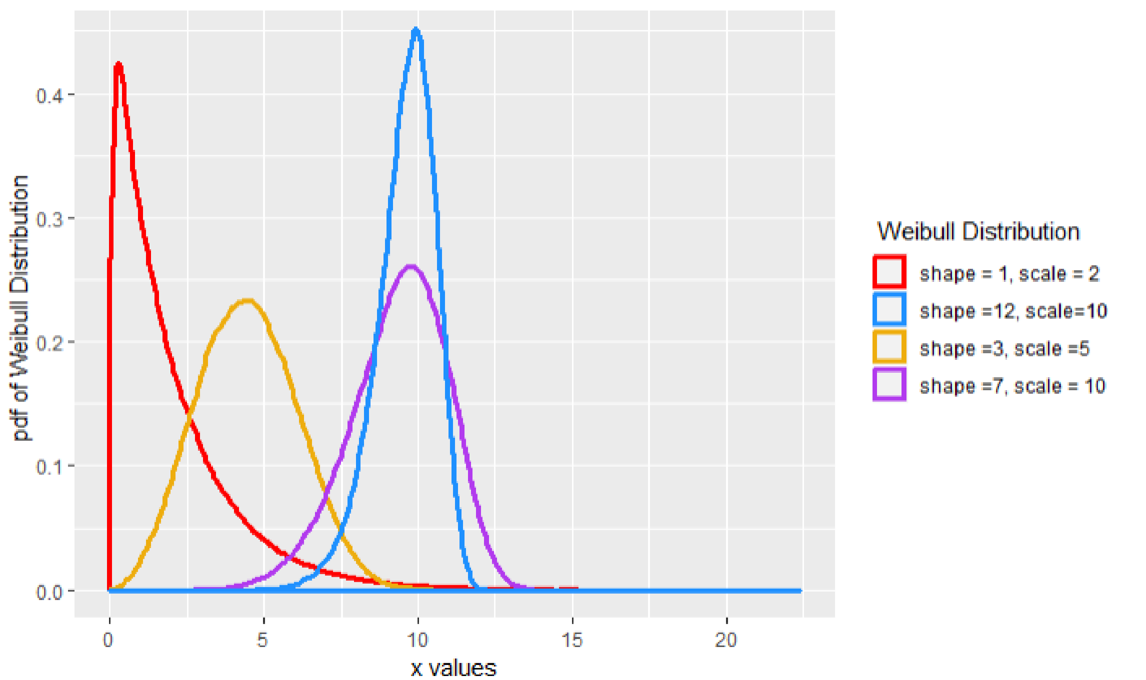
$$\begin{aligned} \frac{d}{d\lambda}(\mathcal{L}(\lambda)) &= \frac{d}{d\lambda}(-n\lambda + \sum_{i=1}^{n} x_i ln(\lambda)) \\ &= -n + \frac{\sum_{i=1}^{n} x_i}{\lambda} \end{aligned} \tag{16}$$

Now, setting = 0 and solving for $\lambda$, we get that

$$\lambda = \frac{\sum_{i=1}^{n} x_i}{n}$$

Note that the second derivative $< 0$. Thus, the MLE of $\hat{\lambda} = \bar{x}$

## 2.4 Weibull Distribution




The Weibull distribution is a very applicable distribution that is used in product engineering to figure out the failure rate or life of a product. The pdf of a Weibull distribution is given by

$$f(x) = \frac{k}{\lambda}(\frac{x}{\lambda})^{k-1}e^{-(\frac{x}{\lambda})^k}$$

where $k \geq 0, \lambda > 0$. The two parameters k and $\lambda$ are the shape and scale parameter respectively. Another interesting property of the Weibull distribution is that when k = 1, the Weibull distribution is the exponential distribution. Also, certain parameters will make the Weibull distribution approximately normal.

The likelihood function is given by

$$L(k,\lambda) = \prod_{i=1}^{n} \frac{k}{\lambda}(\frac{x}{\lambda})^{k-1} e^{-(\frac{x}{\lambda})^k}$$

Thus, the log-likelihood function is given by

$$\begin{aligned}
\mathcal{L}(k,\lambda) &= ln(\prod_{i=1}^{n} \frac{k}{\lambda}(\frac{x}{\lambda})^{k-1} e^{-(\frac{x}{\lambda})^k}) \\
&= nln(\frac{k}{\lambda}) + \sum_{i=1}^{n} ln(x_i^{k-1}) - n(k-1)ln(\lambda) + \sum_{i=1}^{n} \big(\frac{x_i}{\lambda}\big)^k \\
&= nln(k) + (k-1)\sum ln(x_i) - nkln(\lambda) - \lambda^{-k}\sum x_i^k
\end{aligned} \tag{17}$$

Now we must take the derivative with respect to $\lambda$. It follows that,

$$\begin{aligned}
\frac{\partial}{\partial\lambda}(\mathcal{L}(k,\lambda)) &= \frac{\partial}{\partial\lambda}(nln(k) + (k-1)\sum ln(x_i) - nkln(\lambda) - \lambda^{-k}\sum x_i^k) \\
&= \frac{-nk}{\lambda} + k\lambda^{-k-1}\sum x_i^k \\
&= \frac{-nk}{\lambda} + \frac{k}{\lambda^{k+1}}\sum x_i^k
\end{aligned} \tag{18}$$

Now, set equation (18) = 0 and we get

$$\begin{aligned}
0 &= \frac{-nk}{\lambda} + \frac{k}{\lambda^{k+1}}\sum x_i^k \\
&= -nk\lambda^k + k\sum x_i^k
\end{aligned} \tag{19}$$

Thus,

$$\lambda^k = \frac{\sum_{i=1}^{n} x_i}{n}$$

Similarly,

$$
\begin{aligned}
\frac{\partial}{\partial k}(\mathcal{L}(k,\lambda)) &= \frac{\partial}{\partial k}(nln(k) + (k-1)\sum ln(x_i) - nkln(\lambda) - \lambda^{-k}\sum x_i^k) \\
&= \frac{n}{k} - nln(\lambda) + \sum_{i=1}^{n} ln(x_i) - \frac{\partial}{\partial k}\sum_{i=1}^{n}(\frac{x_i}{\lambda})^k \\
&= \frac{n}{k} - nln(\lambda) + \sum_{i=1}^{n} ln(x_i) - \sum_{i=1}^{n}\frac{\partial}{\partial k}(\frac{x_i}{\lambda})^k \\
&= \frac{n}{k} - nln(\lambda) + \sum_{i=1}^{n} ln(x_i) - \sum_{i=1}^{n}(\frac{x_i}{\lambda})^k ln(\frac{x_i}{\lambda}) \\
&= \frac{n}{k} - nln(\lambda) + \sum_{i=1}^{n} ln(x_i) - \frac{1}{\lambda^k}\sum_{i=1}^{n} x_i^k(ln(x_i) - ln(\lambda))
\end{aligned} \tag{20}
$$

We can now substitute for $\lambda^k$ and we get

$$
\begin{aligned}
\frac{n}{k} - nln(\lambda) + \sum_{i=1}^{n} ln(x_i) - \frac{1}{\lambda^k}\sum_{i=1}^{n} x_i^k(ln(x_i) - ln(\lambda)) &= \frac{n}{k} - nln(\lambda) + \sum_{i=1}^{n} ln(x_i) - \frac{n}{\sum x_i^k}\sum_{i=1}^{n} ln(x_i) - x_i^k ln(\lambda)) \\
&= \frac{n}{k} - nln(\lambda) + \sum_{i=1}^{n} ln(x_i) - \frac{n}{\sum x_i^k}\sum_{i=1}^{n}(x_i^k ln(x_i) - x_i^k ln(\lambda)) \\
&= \frac{n}{k} - nln(\lambda) + \sum_{i=1}^{n} ln(x_i) - \frac{n\sum_{i=1}^{n}(x_i^k ln(x_i))}{\sum x_i^k} + nln(\lambda) \\
&= \frac{n}{k} + \sum_{i=1}^{n} ln(x_i) - \frac{n\sum x_i^k ln(x_i)}{\sum x_i^k} = 0
\end{aligned} \tag{21}
$$

Multiplying this by $\frac{-1}{n}$, we get that

$$\frac{n\sum x_i^k ln(x_i)}{\sum x_i^k} - \frac{1}{k} - \frac{1}{n}\sum ln(x_i) = 0$$

The MLE of this cannot be further simplified by hand. We must use statistical software to solve for it.

# 3 Asymptotic Distribution of MLE

Now that we know about the likelihood function, $L(\theta)$, we want to understand more about $L(\hat{\theta})$. We can approximate this function using Taylor series expansion about $\theta$ so that we get

$\frac{\partial[lnL(\theta)]}{\partial\theta} + (\hat{\theta}-\theta)\frac{\partial^2[lnL(\theta)]}{\partial\theta^2} = 0$

Now we can solve for thetas and we get that,

$\hat{\theta}-\theta = \frac{\frac{\partial\mathcal{L}(\theta)}{\partial\theta}}{\frac{\partial^2\mathcal{L}(\theta)}{\partial\theta^2}}$

Note that $\frac{\partial lnL(\theta)}{\partial\theta} = \sum_{i=1}^{n}\frac{\partial[lnf(X_i;\theta)]}{\partial\theta}$ Thus, we can say that $Y_i = \sum_{i=1}^{n}\frac{\partial[lnf(X_i;\theta)]}{\partial\theta}$ and that we now have n independent random variables. Therefore, by CLT, we now have an approximate normal distribution. From this, we can find that the mean in the continuous case is

$$\begin{aligned}\int_{-\infty}^{\infty}\frac{\partial[lnf(x;\theta)]}{\partial\theta}f(x;\theta)dx &= \int_{-\infty}^{\infty}\frac{\partial[f(x;\theta)]}{\partial\theta}\frac{f(x;\theta)}{f(x;\theta)}dx \\ &= \int_{-\infty}^{\infty}\frac{\partial[f(x;\theta)]}{\partial\theta}dx \\ &= \frac{\partial}{\partial\theta}\int_{-\infty}^{\infty}[f(x;\theta)]dx \\ &= \frac{\partial}{\partial\theta}(1) = 0\end{aligned} \tag{22}$$

therefore the mean = 0. Since the mean is $\int_{-\infty}^{\infty}\frac{\partial[lnf(x;\theta)]}{\partial\theta}f(x;\theta)dx$ we can take the derivative with respect to $\theta$ and get that

$$\int_{-\infty}^{\infty}(\frac{\partial[lnf(x;\theta)]}{\partial\theta}f(x;\theta) + \frac{\partial[lnf(x;\theta)]}{\partial\theta}\frac{\partial[f(x;\theta)]}{\partial\theta}) = 0 \tag{23}$$

We can factor out $\frac{\partial[lnf(x;\theta)]}{\partial\theta}$ which leaves us with

$\int_{-\infty}^{\infty}(\frac{\partial^2[lnf(x;\theta)]}{\partial\theta^2}f(x;\theta) + \frac{[f(x;\theta)]}{\partial\theta}) = 0$

and thus,

$$\frac{\partial^2[lnf(x;\theta)]}{\partial\theta^2}f(x;\theta) = \frac{[f(x;\theta)]}{\partial\theta} \tag{24}$$

now substitute and we get that

$$\int_{-\infty}^{\infty}(\frac{\partial[lnf(x;\theta)]}{\partial\theta})^2 = -\int_{-\infty}^{\infty}(\frac{\partial^2[lnf(x;\theta)]}{\partial\theta^2}f(x;\theta)) \tag{25}$$

Since we know that E(Y) = 0, the variance for $Y_1 ... Y_n$ is given by $-nE(\frac{\partial^2[lnf(X;\theta)]}{\partial\theta^2})$. Now manipulate the following equation

$$\hat{\theta} - \theta = \frac{\frac{\partial \mathcal{L}(\theta)}{\partial \theta}}{\frac{\partial^2 \mathcal{L}(\theta)}{\partial \theta^2}}$$

and we get that

$$\frac{\sqrt{n}(\hat{\theta} - \theta)}{\frac{1}{\sqrt{-E(\partial^2(ln(f(X;\theta)/\partial\theta^2))}}} = \frac{\frac{\partial(\mathcal{L}(\theta))/\partial\theta}{\sqrt{-nE(\partial^2(ln(f(X;\theta)/\partial\theta^2))}}}{\frac{\frac{-1}{n}\frac{\partial^2(\mathcal{L}(\theta))}{\partial\theta^2}}{-E(\partial^2(ln(f(X;\theta)/\partial\theta^2)}} \tag{26}$$

so that the right hand side of the equation is the sum of n independent random variables. Thus, by CLT, the distribution of $\hat{\theta}$ is approximately a N(0,1) distribution with mean = $\theta$ and standard deviation

$$\frac{1}{\sqrt{-nE(\partial^2(lnf(X;\theta)))/\partial\theta^2}}$$

## 3.1 Examples of Asymptotic Distribution of MLE

Now let us do some examples.

Note that the MLE of a normal distribution are $\mu = \bar{x}$ and $\sigma = \sigma^2$. We also know that the $lnf(x;\theta) = -\frac{1}{2}ln(2\pi\sigma^2) - \frac{(x-\theta)^2}{2\sigma^2}$

$$\begin{aligned}\frac{\partial}{\partial\theta}(lnf(x;\theta)) &= \frac{\partial}{\partial\theta}(-\frac{1}{2}ln(2\pi\sigma^2) - \frac{(x-\theta)^2}{2\sigma^2}) \\ &= \frac{(x-\theta)}{\sigma^2}\end{aligned} \tag{27}$$

Now taking the second derivative with respect to theta , we get

$$\begin{aligned}\frac{\partial^2}{\partial\theta^2}(lnf(x;\theta)) &= \frac{\partial^2}{\partial\theta^2}(\frac{(x-\theta)}{\sigma^2}) \\ &= \frac{1}{\sigma^2}\end{aligned} \tag{28}$$

Now, substitute into the standard deviation formula,$\frac{1}{\sqrt{-nE(\partial^2(lnf(X;\theta)))/\partial\theta^2}}$. Thus, the mean = $\theta$ and standard deviation = $\frac{\sigma^2}{n}$

Now let us look at another example with the exponential distribution. In this example we will solve for expectation lastly, to show that order doesn't matter. Note that $\hat{\theta} = \bar{x}$ and $\ln f(x;\theta) = \ln\theta + (\theta - 1)\ln x$. Now taking the 1st and 2nd derivative we get

$$\frac{dlnf(x;\theta)}{d\theta} = \frac{-1}{\theta} + \frac{x}{\theta^2}$$
$$\frac{d^2\theta lnf(x;\theta)}{d\theta^2} = \frac{1}{\sigma^2} - \frac{2x}{\sigma^3}$$

now solve for the expectation

$$-E[\frac{1}{\theta^2} - \frac{2X}{\theta^3}] = -\frac{1}{\theta^2} + \frac{2\theta}{\theta^3} = \frac{1}{\theta^2}$$

$$\frac{1}{\sqrt{n\frac{1}{\theta^2}}} = \frac{1}{\frac{\sqrt{n}}{\theta}} = \frac{\theta}{\sqrt{n}}$$

Thus $\bar{x}$ has an approximate normal distribution with mean $\theta$ and standard deviation $\frac{\theta}{\sqrt{n}}$

Let us do one final example using the Poisson Distribution. Note that the MLE of the Poisson distribution is $\hat{\lambda} = \bar{x}$ and that $lnf(x;\lambda) = xln\lambda - \lambda - ln(x!)$. Thus the 1st and 2nd derivative are

$$\frac{dlnf(x;\lambda)}{d\lambda} = \frac{x}{\lambda} - 1$$
$$\frac{d^2lnf(x;\lambda)}{d\lambda^2} = \frac{-x}{\lambda^2}$$

Thus, the expectation is

$$E(-\frac{x}{\lambda}) = \frac{\lambda}{\lambda^2} = \frac{1}{\lambda}$$

Now, we can solve for the standard deviation to get that standard deviation = $\sqrt{\frac{\lambda}{n}}$ Thus, $\hat{\lambda}$ has an approximate normal distribution with mean, $\lambda$, and standard deviation, $\sqrt{\frac{\lambda}{n}}$.

# 4 Methods and Application

## 4.1 Motivation

Oftentimes we are sitting in class and we wonder "is this useful to us in the real-world?" For this research, I wanted to show a first hand application of MLE that would be meaningful to my university and all the college students within it. Therefore, I focused on the common question: "Is there enough time to eat without being late to my next class?" Often in college, students and professors alike find it difficult to eat at normal times due to scheduling conflicts. Thus, if I could find the average time it would take you to grab food from the Chick-fil-A and Subway, it would be beneficial on a personal level, which led me to observing the average service time for both these restaurants. I further wanted to see the application of MLE on a larger scale and therefore observed the inter-arrival time of the customers and counted the number of people arriving during that hour to both of these locations.

By knowing the inter-arrival time and number of people showing up daily, businesses could figure out how many workers they should have on a certain day and at what times of day they should bring in more staff. This would help businesses become more efficient because they could make sure the customers have shorter service times and they would also waste less food daily.

## 4.2 Data Collection

My study took place on weekdays from Monday, October 15th until Friday, November 9th from 12:00 pm -1pm in the Bobcat Food Court. I chose that interval of time specifically because it is often viewed as the lunch hour. While there, I kept track of the service time, inter-arrival time, and number of people arriving during that hour.Because I wanted the results of this study to help those who would be in a rush to get food, I focused on several factors. First, I omitted customers who bought just a drink or milkshake since I primarily wanted to focus on customers that ordered food. Second, I specifically waited to begin my count until the person walked into the line since many students spent time talking with friends or hanging out before they got in line. Once in the line, service time was counted in seconds until the moment they had their food in their hand. Inter-arrival time is also measured in seconds and is the amount of time between two customers getting into the line for any given restaurant. Number of people during the hour was a count variable and is simply the number of people that arrived to each restaurant daily from 12pm - 1pm. All of my data was then input into RStudio for analysis. Below is the header for my data

| Ctime | Stime | CIAtime | SIAtime | Cdaily | Sdaily |
|---|---|---|---|---|---|
| 490 | 447 | 4 | 13 | 82 | 29 |
| 463 | 388 | 24 | 49 | 71 | 26 |
| 449 | 272 | 49 | 0 | 69 | 33 |
| 488 | 241 | 30 | 113 | 91 | 23 |
| 607 | 188 | 21 | 20 | 63 | 22 |
| 581 | 255 | 93 | 28 | 58 | 29 |

Table 1: Header of Data

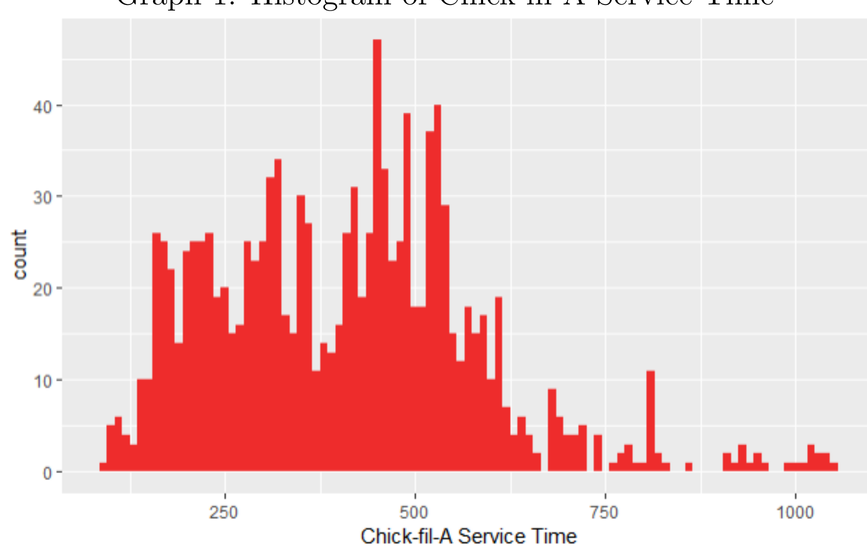
```
data.frame':   1194 obs. of  6 variables:
 $ Ctime   : int  490 463 449 488 607 581 725 741 714 655 ...
 $ Stime   : int  447 388 272 241 188 255 287 337 248 279 ...
 $ CIAtime: int  4 24 49 30 21 93 20 39 40 0 ...
 $ SIAtime: int  13 49 0 113 20 28 12 61 80 0 ...
 $ Cdaily : int  82 71 69 91 63 58 51 51 63 72 ...
 $ Sdaily : int  29 26 33 23 22 29 19 28 34 45 ...
     Ctime            Stime           CIAtime          SIAtime
 Min.   :  86.0   Min.   :115.0   Min.   :  0.00   Min.   :  0.00
 1st Qu.: 277.5   1st Qu.:232.5   1st Qu.: 17.00   1st Qu.: 51.75
 Median : 412.0   Median :326.0   Median : 44.00   Median :100.50
 Mean   : 407.7   Mean   :352.3   Mean   : 58.29   Mean   :132.26
 3rd Qu.: 518.0   3rd Qu.:444.5   3rd Qu.: 77.00   3rd Qu.:181.50
 Max.   :1052.0   Max.   :918.0   Max.   :722.00   Max.   :896.00
                  NA's   :683     NA's   :20       NA's   :702
     Cdaily          Sdaily
 Min.   :39.00   Min.   :15.00
 1st Qu.:50.75   1st Qu.:23.00
 Median :59.00   Median :29.00
 Mean   :59.70   Mean   :27.75
 3rd Qu.:66.75   3rd Qu.:31.25
 Max.   :91.00   Max.   :45.00
 NA's   :1174    NA's   :1174
```

Table 2: Summary of Data

Overall, there were 1194 observations for service time at Chick-fil-A and 511 observations for service time at Subway. The shortest service time for Chick-fil-A was 1 minute and 26 seconds and the longest was 17 minutes and 32 seconds; Subway on the other hand, ranged from 1 minute and 55 seconds to 15 minutes and 18 seconds.

# 5 Results

Graph 1: Histogram of Chick-fil-A Service Time

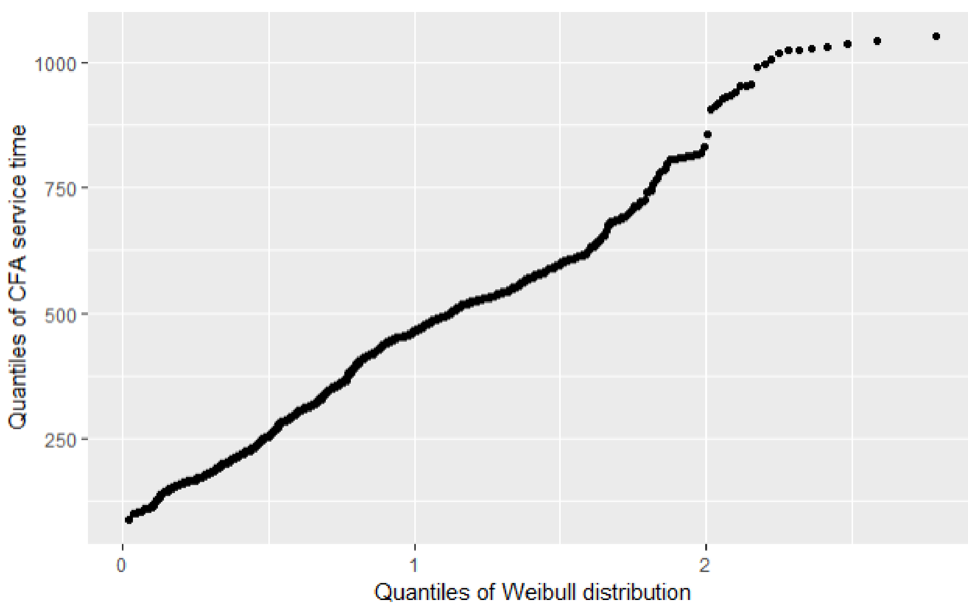



After observing the histogram, we wanted to verify if it was a normal distribution or not. Below is the summary for the tests for normality

```
        One-sample Kolmogorov-Smirnov test

data:  Ctime
D = 1, p-value < 2.2e-16
alternative hypothesis: two-sided

        Anderson-Darling normality test

data:  Ctime
A = 7.0472, p-value < 2.2e-16

        Shapiro-Wilk normality test

data:  Ctime
W = 0.96098, p-value < 2.2e-16
```

We can see that the p-value is approximately 0; thus we can conclude that this distribution is not a normal distribution. After some tests, we came to the conclusion that Chick-fil-A service time followed a Weibull distribution. Once the Weibull distribution was verified, we wanted to see how close our data matched Weibull distribution, so we compared the quantiles of Chick-fil-A service time and the quantiles of the Weibull distribution.

Graph 2: Quantiles of CFA Service Times vs Quantiles of Weibull Distribution

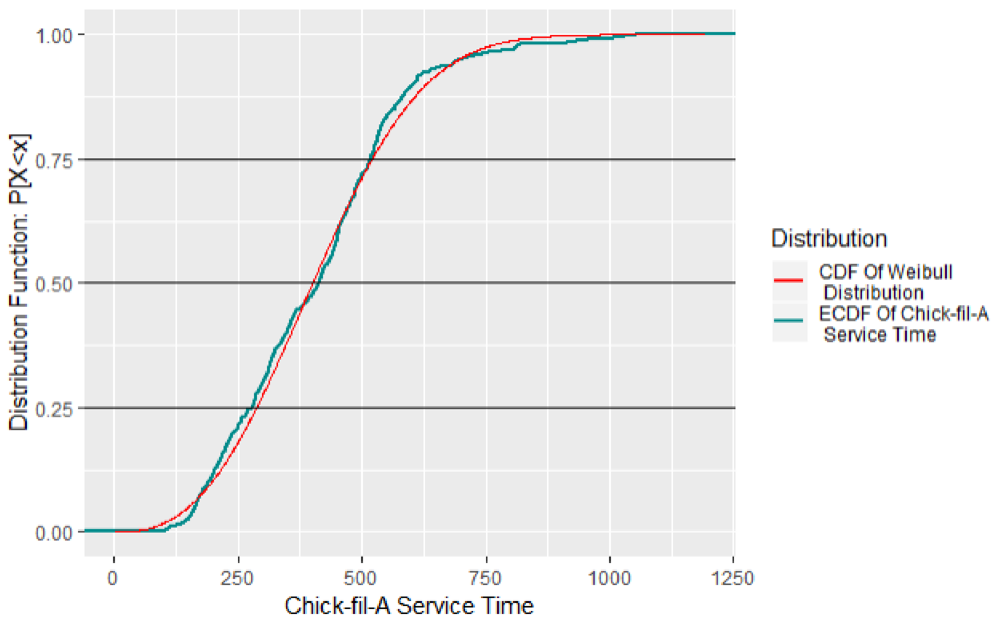



From Graph 2, we can see that there is a linear relationship between the quantiles; since it is a very nice linear line, we can conclude that the Weibull distribution is a very good model for Chick-fil-A service time.

We then graphed the CDF of the Weibull Distribution against the ECDF of Chick-fil-A service time to get a more visual representation of how closely they are related. The ECDF (Empirical cumulative distribution function) is given by

$$G(t) = \frac{1}{n}\sum_{i=1}^{n} 1_{x_{i \leq t}}$$

Graph 3: Weibull CDF vs ECDF of CFA Service Times

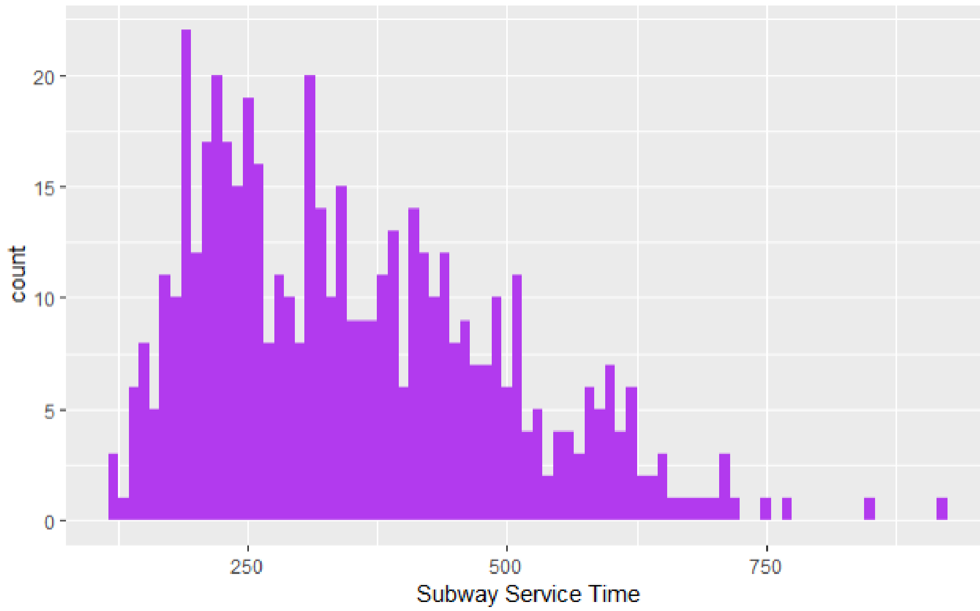



From Graph 3, we can see how closely the functions align. Then, using R, we find the MLE of the function and find that the best cdf of Chick-fil-A service time is

$$F(t) = 1 - e^{-\left(\frac{t}{460}\right)^{2.65}}$$

Repeating this same process with the Subway data, we find the following results for Subway service time.

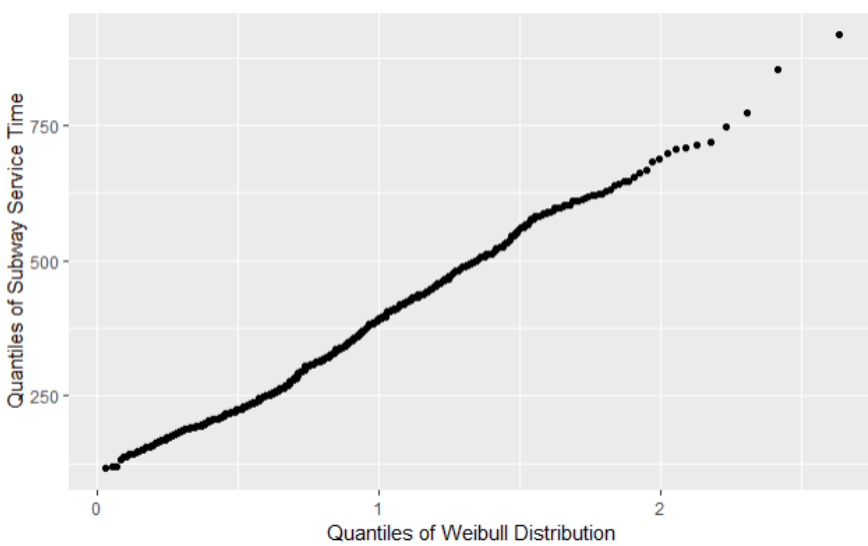
Graph 4: Histogram of Subway Service Time



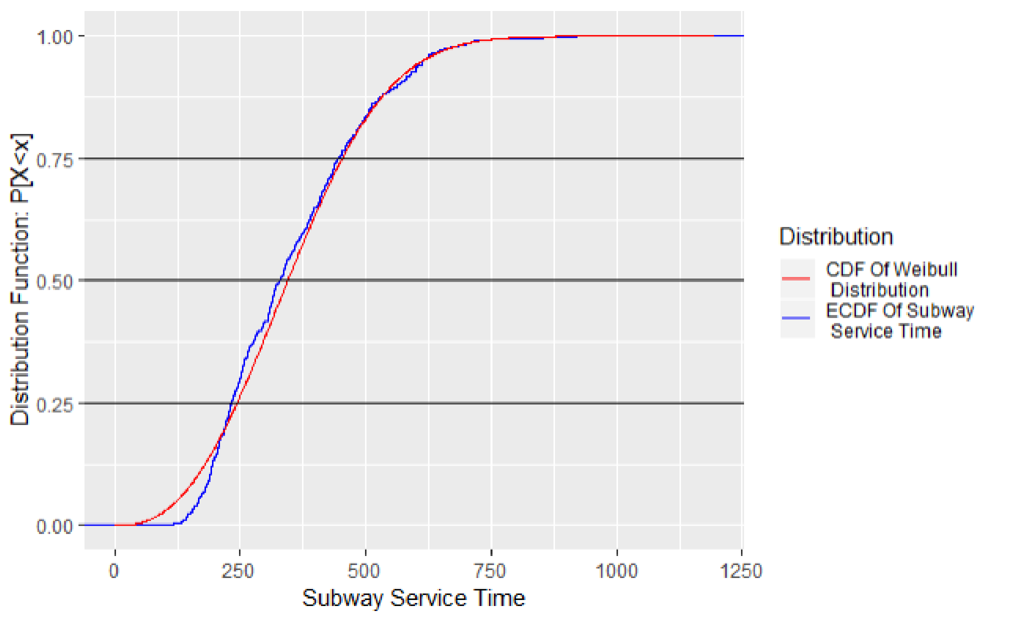

```
Shapiro-Wilk normality test

data:  Stime
W = 0.95322, p-value = 1.204e-11
```

After running the same tests as we did on the Chick-fil-A data, we find that Subway service time is best represented as a Weibull distribution.

Graph 5: Quantiles of Weibull vs Quantiles of Subway



Graph 6: CDF of Weibuill Distribution vs ECDF of Subway Service Time

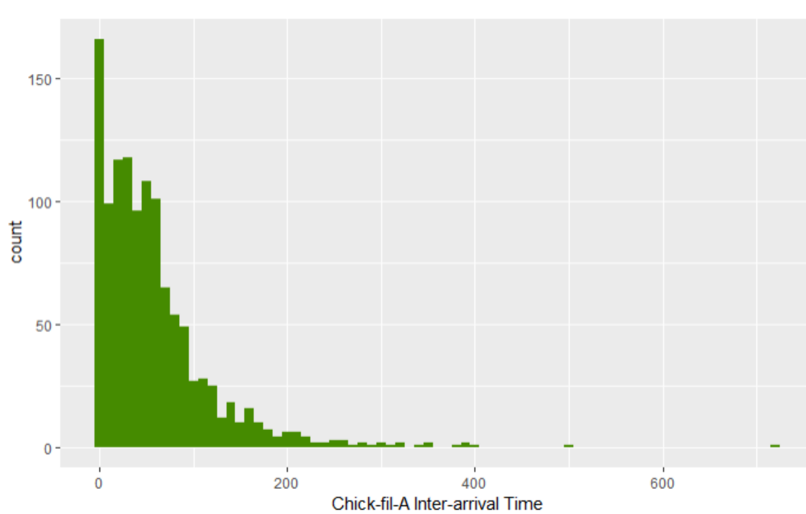



Now we use R to calculate the MLE and we get that the best estimate of the cdf of Subway service time is given by

$$F(t) = 1 - e^{-\left(\frac{t}{398}\right)^{2.62}}$$

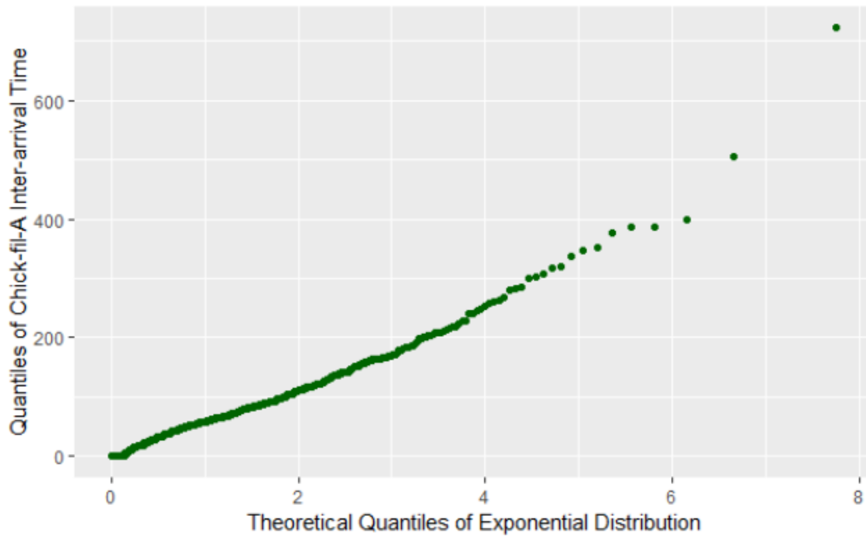
Next, we plot a Histogram of Chick-fil-A inter-arrival time

Graph 7: Histogram for CFA Inter-arrival Time

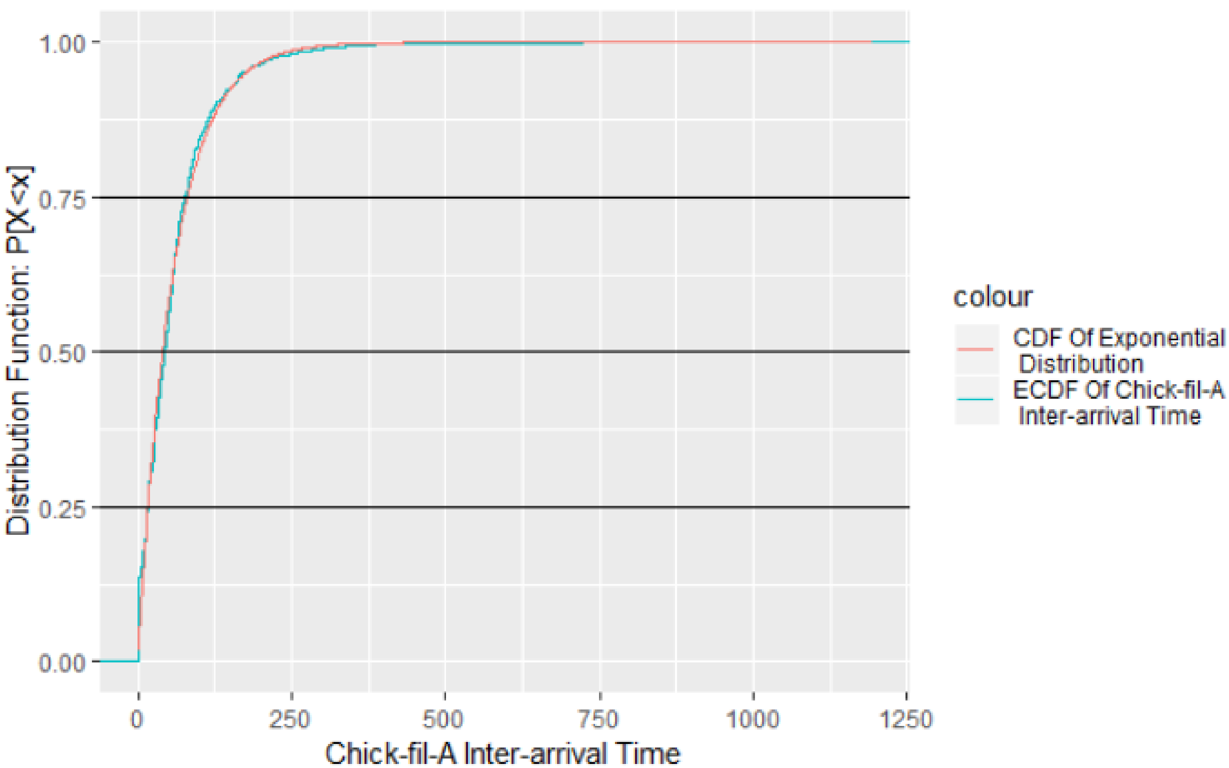



We then find the best model for Chick-fil-A inter-arrival time to be an exponential distribution, which is expected after observing Graph 7.

Graph 8: Quantiles of CFA Inter-arrival vs Quantiles of Exp. Distribution




Graph 9: CDF of Exp. Distribution vs ECDF of Subway Inter-arrival Time

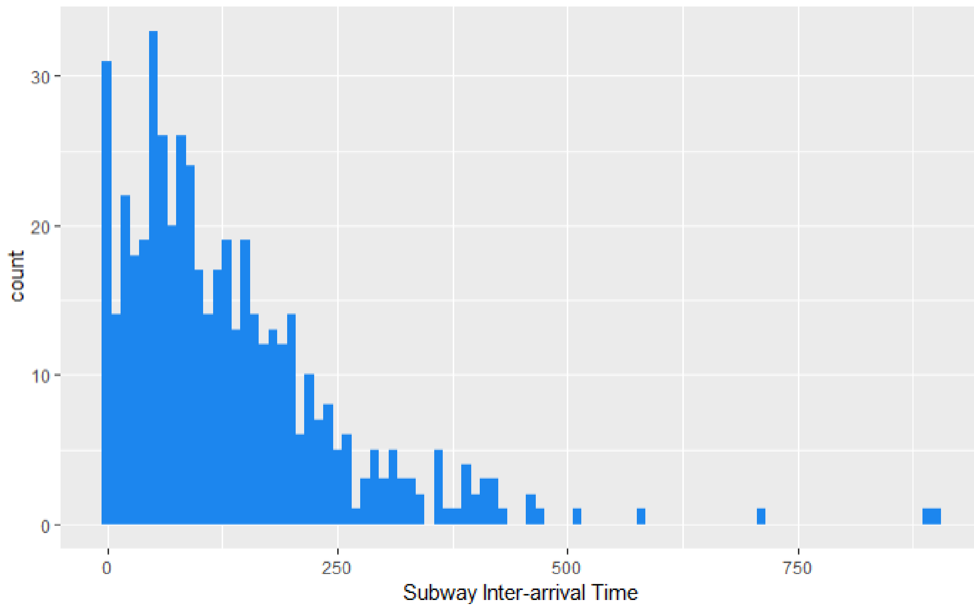



Besides a few outliers, Graph 8 has a strong linear relationship; if we look at Graph 9 the lines are basically identical, which further reinforces that this is a great distribution for our data. Thus we can conclude that the Chick-fil-A inter-arrrival time follows an exponential distribution with cdf

$$F(t) = 1 - e^{-.0076t}$$

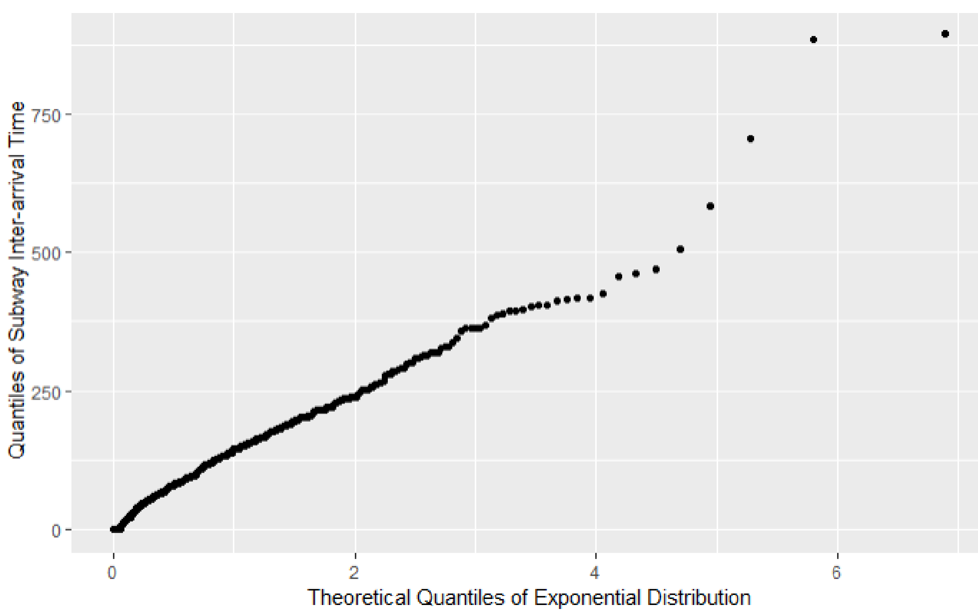
Now we run the same procedure for Subway inter-arrival time and we produce the following plots

Graph 10: Histogram of Subway Inter-arrival Time

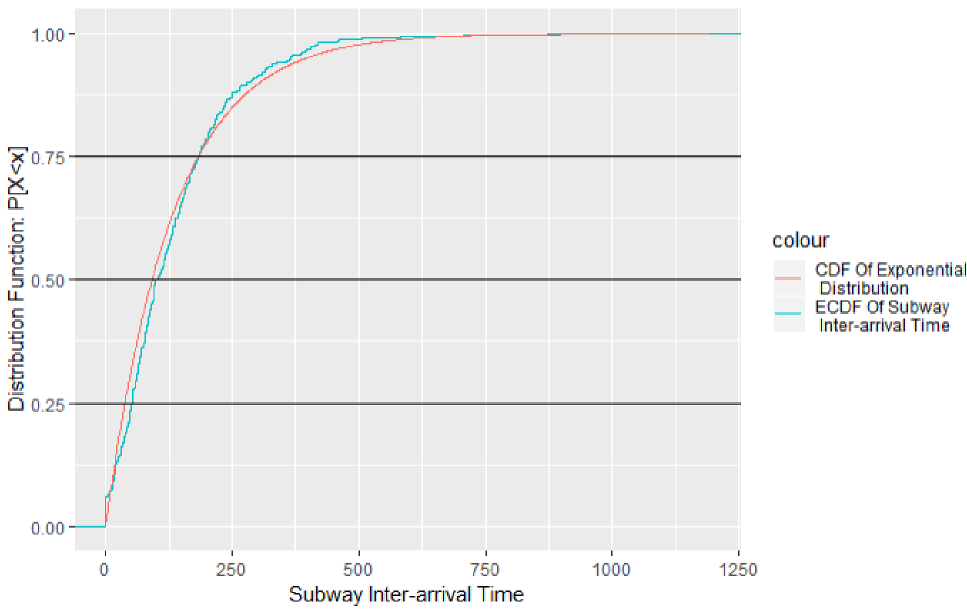



Similar to Chick-fil-A inter-arrival time, Subway inter-arrival time is best modeled by an exponential distribution.

Graph 11: Quantiles of Subway inter-arrival Times vs Quantiles of the Exp. Distribution



Graph 12: CDF of Exp. Distribution vs ECDF of Subway Inter-arrival Time




Besides a few outliers, Graph 11 has a strong linear relationship and the curves on Graph 12 are also very similar, We can now calculate the MLE to find that the best cdf for Subway inter-arrival time is given by

$$F(t) = 1 - e^{-.0076t}$$

Since we only sampled for 20 days, we did not have enough data to find out about the distribution for the number of people showing up during the hour. Given the fact that the number of people showing up during the hour is a count variable, we assume that, if we have enough data, then it would have followed a Poisson distribution.

- CFA service time cdf :$F(t) = 1 - e^{-\left(\frac{t}{460}\right)^{2.65}}$
- Subway service time cdf:$F(t) = 1 - e^{-\left(\frac{t}{398}\right)^{2.62}}$
- CFA inter-arrival time cdf:$F(t) = 1 - e^{-.0172t}$
- Subway inter-arrival time cdf: $F(t) = 1 - e^{-.0076t}$

Since we have the cdf distributions, we can now use these to estimate the probability of waiting x or more minutes from 12pm-1pm on any weekday.

What is the probability that you wait 5 or more minutes at Chick-fil-A?

$P[X \geq 300] = e^{-\left(\frac{t}{460}\right)^{2.65}} = e^{-\left(\frac{300}{460}\right)^{2.65}} = .7245$

It is 72.45% likely that you will wait more than 5 minutes at Chick-fil-A.

How about for 10 minutes or more?

$P[X \geq 600] = e^{-\left(\frac{t}{460}\right)^{2.65}} = e^{-\left(\frac{600}{460}\right)^{2.65}} = .1324$

It is 13.24% likely that you will wait more than 10 minutes at Chick-fil-A.

Now let us do the same calculations for Subway.

What is the probability that you wait 5 or more minutes at Subway?

$P[X \geq 300] = e^{-\left(\frac{t}{398}\right)^{2.62}} = e^{-\left(\frac{300}{398}\right)^{2.62}} = .6207$

It is 62.07% likely that you wait more than 5 minutes at Subway.

More than 10 minutes?

$P[X \geq 600] = e^{-\left(\frac{t}{398}\right)^{2.62}} = e^{-\left(\frac{600}{398}\right)^{2.62}} = .0533$

It is 5.33% likely that you wait more than 10 minutes at Subway.

What is the probability that no one will come to Chick-fil-A for 5 minutes or more?

$P[X \geq 300] = e^{-.0172(300)} = .0057$

There is a .57% chance that no one will come to Chick-fil-A for 5 minutes or more.

What is the probability that no one will come to Subway for 5 minutes or more?

$P[X \geq 300] = e^{-.0076(300)} = .1023$

There is a 10.23% chance that no one shows up to Subway for 5 minutes or more.

We found that the average wait time is 6 minutes and 48 seconds for Chick-fil-A and 5 minutes and 52 seconds for Subway. Also, after running a T-test, we discovered that service time for CFA is significantly longer than service time for Subway. We can also say with 95% confidence that Chick-fil-A service time will take 39.5-71.3 seconds longer on average than Subway service time.

# 6 Further Research

The first thing to do for any sample would be to collect more data. If we were able to collect more data, then we could also estimate the number of people that would show up during the given hour. We could expand this to different times of the day and figure out the time distribution for any given hour. Also, since our sample is not extremely large, we cannot separate the time it would take based on the day, which would be more accurate. Another thing I would find interesting is to conduct a study similar to this on an off-campus restaurant and compare the results to the online service time histogram provided from Google.

# References

Hogg. Robert V. Tanis. Elliot. Zimmerman. Dale. Probability and Statistical inference. ninth ed.. Pearson. 2015.

Wackerly, Dennis D., et al. Mathematical Statistics with Applications. seventh ed., Duxbury Press., 2008.