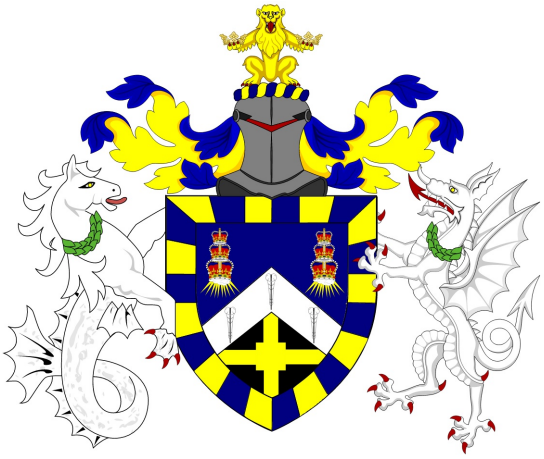
Mathematics MSc Dissertation MMTHM038, 2020/21

# The German Tank Problem


## Koumari Gontla, ID 200282280

Supervisor: Dr Wolfram Just

A thesis presented for the degree of

Master of Science in *Mathematics*


School of Mathematical Sciences

Queen Mary University of London

# Declaration of original work

This declaration is made on September 5, 2021.

**Student's Declaration:** I Koumari Gontla hereby declare that the work in this thesis is my original work. I have not copied from any other students' work, work of mine submitted elsewhere, or from any other sources except where due reference or acknowledgement is made explicitly in the text, nor has any part been written for me by another person.

Referenced text has been flagged by:

1. Using italic fonts, **and**
2. using quotation marks "...", **and**
3. explicitly mentioning the source in the text.

# Abstract

The German Tank problem is a famous problem that first appeared during World War 2. The allies tried to predict the number of tanks Germany was producing with mainly two techniques. One using Frequentist statistics and the other being Bayesian statistics. They were able to predict the number of tanks being produced based on the serial numbers of the captured tanks.

We will generate random samples of numbers (tank serial numbers) to mimic the method of capturing random tank pieces. We will use fixed dummy numbers for our variables such as: sample size and population to find a point estimate of the true population size with certain formulas. More specifically, the formulas used to find point estimates for the population size for both the Frequentist method and the Bayesian method. This involves finding the pmf and using it to derive the expectation and variance.

We will conduct our own experiments to see how this method is done and get a better understanding of it. We will see if we can use other methods or alterations of known methods to get a better estimate. Such as using a different prior in the Bayesian method in R and deriving a value with computations. We will see in our results which approach gives more accurate results.

# Chapter 1

# Introduction

## 1.1 Historical Background

During World War 2 the allies wanted to find out the production of German tanks. Before, the US and Britain were thought to have superior tanks compared to Germany's Panzer tanks. But new tanks were being produced, the Mark IV and V tanks. They were worried about the new version's capabilities and because of this more information about these tanks were needed. This could be very advantages to have as you can be more prepared on the battlefield. This would later come to use in the Western Front. Thus, to do this they adopted 2 methods: conventional intelligence and atatistical intelligence. In some ways conventional intelligence was used in conjunction with statistical methods. For example, this was done for the estimation of the Panther Tank production.

However this was different, the first method used was conventional intelligence. It was a more conventional route where they tried to ascertain the number of tanks by methods such as interrogation, intercepting encoded messages, espionage, and tips etc. This proved to be inefficient as they found the estimating numbers to be incorrect and unreliable. Therefore, the next course of action was to ask statistical intelligence to see if they could statistically

estimate the number of tanks being produced.

## 1.2 Motivation

The motivation for this project is to find the most effective method to estimate population size as accurately as we can. We analyse and use the methods taken by statistical intelligence as the numbers they obtained were much closer to the actual figures. Statistics estimated that Germany was producing around 256 tanks per month. This was later confirmed to be exactly the number as recorded from the Ministry of Albert Speer. Thus, this was seen as a much better method of estimation.
The approach used was to use the unique serial numbers imprinted onto the captured tanks. The serial numbers were useful to attach onto the tanks to keep track of the production levels for Germany and in turn useful for the allies to predict it. On the battlefield there would often be broken and destroyed parts left over. The allies captured these to analyse and study the manufacturing capabilities. In this case they could be used to infer the number of produced tanks.
We have a chart comparing estimates from statistics and intelligence to the actual German records for tank produced.

Figure 1.1: Chart of German Records

| Date | Statistical Estimate | Intelligence Estimate | German Records |
|---|---|---|---|
| June 1940 | 169 | 1000 | 122 |
| June 1941 | 244 | 1550 | 271 |
| August 1942 | 327 | 1550 | 342 |

Statistical estimates are significantly much more accurate than intelligence. On average the statistical estimates are off by around 30. However, intelligence estimates tends to be off by over a 1000. We look at point estimates to get a statistical value that should give us a better understanding of the population and thus the true population value. Point estimates can be in

the form of the expectation where it takes an average of all readings. As well as the form of the mode, like the case of maximum likelihood estimate. The variance we obtain should show us how accurate we believe our point estimate to be.

## 1.3 Definitions

- N – Population size parameter
- m – Sample size
- $X_m$ – random variable for the maximum value in a sample
- $x_m$ – maximum value in a sample (sample max)
- $x_i$ – random variable for values in position for i= 1,2,. . . ,m
- l – maximum value in a sample ($x_m$)

### 1.3.1 Example of Method

In this example we will demonstrate how the serial numbers are used on a smaller scale. Say we have a total populations size $N$ of tanks (we are not sure what $N$ is). We collect $m = 5$ samples of destroyed pieces with unique serial numbers and we place them in order $x_i$ . We sample without replacement since each tank will have a unique serial number and we try to find the total.

$$x_1, x_2, x_3, x_4, x_5 = 42, 52, 23, 19, 60$$

Our sample max $x_m = 60$ (and l=60). Note, that since we are trying to find the population estimate it would be the bigger than or equal to the biggest serial number we find in our sample. Our population estimate cannot be lower than our sample size $m$, it can only be exactly equal to $m$ or bigger. However, if we say the number of total tanks produces is our max sample number that

itself is too conservative as it would assume we have captured the most recent tank produced. Thus, the real population size is probably over the sample max. So the sample max is crucial in finding the real population size. We can say the tanks we captured follow a stochastic process. A stochastic process is defined as a collection of observations taken at a specific time. The outcome (observed values) are random variables taken at each time. In our case the random variables are the serial numbers on the captured tanks.

## 1.4 Estimator

An estimator is a statistic that tell you something about a specific population. We will specifically use point estimates as they are single values that give an estimate of a parameter of a certain population. We generate point estimates by taking a sample of the true population we want to find out more about and applying different methods to get them.

### 1.4.1 Example of Sampling for an Estimator

We might want to see what the average age of people is who attend swimming lessons at the gym. We cannot contact all the gym in all the UK so what we would do is contact a select few for our sample, say 20 gyms. We would get this sample and calculate the average age of people $\hat{x}$ in all the samples which is a point estimate of the true population mean (which represents the actual average age $X$ of people who attend swimming lessons in the UK).

## 1.5 Properties of Point Estimate

Point estimators will not be perfect as they are an estimate. Therefore, you have a Bias which is calculated by the difference of the expected value of the estimator and the actual value of the parameter being estimated. Hence if

there is no difference between the expected and actual value the estimator will be unbiased. The less biased an estimator is, the more accurate we would expect it to be.
We also want our point estimate to be consistent. This means that as our sample size increases the point estimate gets closer and closer to the true value of the parameter. To get an accurate estimate, we will need a big sample size. If we want a point estimate to be considered it should move closer to the true value of the parameter as we increase its sample size.
One would define the most efficient estimate to be one that has very low variance, that is unbiased and is consistent.

## 1.6 Variance

Variance is a statistical measurement that refers to the spread between the numbers in a given data set. Specifically, it measure the difference from each number in the set to the set mean i.e from every other number in the set.
In our experiment it would be preferable to get a small variance. This would mean the numbers in the sets are not that far away from each other. This would indicate a more accurate result because of low volatility.
One way we can try and decrease the variance is by increasing the sample size. If we analyse the formula for variance, we would represent the variance as $\sigma^2$, the $i^{th}$ data points as $x_i$ ,mean of data points as $\bar{x}$ and n as the total number of data points.

$$\sigma^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}$$

We can reduce the variance generally by increase the number of samples and thus the data points. This would give us a bigger numerator thus we would hope the result gets smaller. The standard deviation $\sigma$ is almost the same as variance but is expressed as the square root of the variance.

## 1.7 Probability Mass Function (pmf)

The probability mass function is a statistical expression which defines a probability distribution for a discrete random variable. We define this by the probability that our maximum sample value $X_m$ is equal to some value $l$, represented as $P(X_m = l)$ . So for any value of $l$ we can work out the probability it is the maximum. We assume the sample we collect of tank serial numbers follows that of a Uniform Distribution. This value $l$ has uniform distribution such that $N >= l >= m$ since the max serial number has to be bigger than or equal to the sample size collected as well as smaller than or equal to the true maximum serial number.

Thus, for our pmf we want a subset of sample space so that $l = max(x_i)$ (this will be our numerator). Over total sample space (our denominator). We work out the numerator by using the sample size $m$ and the max serial number $l$. Say we have sets of $m$ samples which have the max sample number $l$. That would leave $m - 1$ tanks left (excluding the max $l$). This remaining $m - 1$ tanks would have been selected from all the number below our sample max so $l - 1$. Hence our subset of sample space such that $l = maxx_i$ is equal to $l - 1$ choose $m - 1$ (without replacement and order does not matter).

Similarly, the probability space is the amount of sample sets we can pick from the total population $N$. So we would do $N$ choose $m$ as our denominator.

$$P(X_m = l) = \frac{\binom{l-1}{m-1}}{\binom{N}{m}}$$

Intuitively we can tell that the pmf $P(X_m = l)$, for a value below $m$ that is below the sample size the pmf will have probability 0. This is because any value less than the sample size will not be the maximum serial number sample maximum therefore cannot be the highest possible serial number. As well as for value that is greater than $N$ the total tanks, the pmf will also have probability 0. This is because we cannot have serial number bigger than the highest.

## 1.8 Intro to Frequentist Approach

The approach has a lot to do with frequencies and probabilities being relative frequencies. The frequency of an event is the amount of times it is observed. We will use this method to form histograms of our experiments for this paper.

We will look at the point estimate and maximums of our samples. We look at the unknown parameter $N$ which will we fix. Our $N$ represents our Population size, i.e the total number of tanks that have been made. As well as our sample size $m$ which we will fix. This will be the number of broken parts collected in battle. We derive our expectation and variance of our point estimate and maximum sample value by starting with the probability mass function.

This pmf gives the probability that some discrete random variable $X_m$ is equal to some value $N$. We do this using a binomial coefficient "choose". We want to find the probability that our discrete random variable $X_m = l$ where the event containing $m-1$ elements are chosen from $1, ..., m-1$.

From the pmf we derive the expectation for sample max and re arranging the expectation of sample max to get $n$ (population size) and substituting $E(x_m) = x_m$. Thus we have,

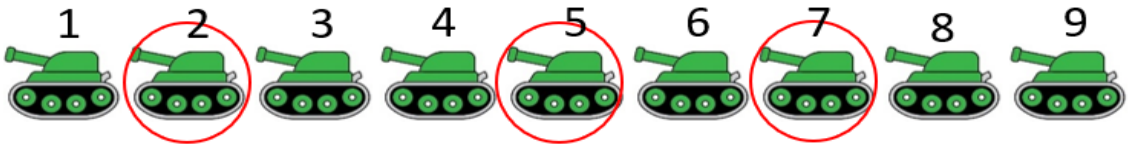
$$n = x_m + (\frac{x_m}{m} - 1)$$

The point estimate we get is essentially our sample max we observe added to $(\frac{x_m}{m} - 1)$. This gives the certainty that we will choose at least the most recently produced tank (sample max $x_m$) that we know of in order to get an accurate estimate. Given that $m < n$, the part in the brackets can be understood intuitively. Our point estimate gets larger as we increase the sample max by $x_m/m$ .

The formula can be understood as the highest serial number we collect plus the number of unobserved tanks. The unobserved tanks make up the intervals

of the serial numbers collected. Let say we order all the tanks from 1 to n and the tanks we capture have equal interval in between their serial numbers. Those equal interval (gaps in between) make up what the next tank we collect could be thus predicting the future highest serial number and the total number of tanks.

### 1.8.1 Example of Frequentist Theory




Lets say we have a population on n=9 tank and each tank has a serial number. We sample m=3 tanks (circled in red). Our sample max is the 7th tank so $x_m = 7$. So our equation $(\frac{x_m}{m} - 1)$=7/3-1=1.3 which we round to 1. This gives us the approximate unobserved tanks in between.

## 1.9 Intro to Bayesian Approach

This approach uses Bayes Theorem. Say we have 2 events, event A and B. The conditional probability of A given B is true can be expressed as:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} = \frac{P(B|A)P(A)}{\sum P(B|A)P(A)}$$

Here our event A is a proposition. For example, it is the proposition that a coin will land on heads 50% percent of the time. Our event B will give new evidence of data. One of the biggest differences between Bayesian and Frequentist is that in Bayesian we have a prior probability. P(A) is a prior probability of A which expresses a prior belief. $P(B|A)$ is our likelihood function,

probability of evidence B given A is true. The second part of the equations is a consequence of the Law of Total Probability $P(B) = \sum P(B|A)P(A)$, we have the same as the numerator with an added summation.
Normally event B is fixed so we want to see the effects of its having been observe for our various possible events A. This is why in Bayesian statistics you will often see $P(A|B) \propto P(A)P(B|A)$. Thus the posterior distributions is directly proportional to out prior multiplied by likelihood.

## 1.10 Maximum Likelihood Estimation (MLE)

The MLE is a method of estimating the parameters of a statistical model given some observations. We do this by finding the values of the parameters which will maximise the likelihood of making the observations given the parameters. Our first Bayesian method we assume a Uniform prior and second a Poisson.
Our MLE can be seen as a special case of the maximum a posterior (MAP) estimation. The MAP is like MLE in that it is the mode of the of the posterior distribution. We use MAP to find a point estimate of a quantity (unobserved) on the basis of data. The difference between MLE and MAP is the use or need of a prior distribution. We will explain briefly the MLE method. We want to use the MLE method to find information about the population with the sample we collect from an unknown sample distribution. It is finding the optimum way to fit a distribution to the data.
Specifically, we want the joint probability distribution of the random variables $y_1, y_2, ...$ (not necessarily independent and identically distributed). There are unique vectors $\theta = [\theta_1, \theta_2, ..., \theta_k]^T$ of parameters that indicates the probability distributions within a parametric family. A parametric family is a family of objects which are related and the differences depend on the select chosen values for a set of parameters. In this case a parametric family $f(;\theta)|\theta \in \Theta)$ where the parameter space is $\Theta$ and represents a finite space.

For the observed data sample $y_1, y_2, .., y_n$. We can analyse the joint density and that gives us a real value function:

$$L_n(\theta) = L_n(\theta; y) = f_n(y; \theta)$$

This is our likelihood function. Where the $f_n(y; \theta)$ is a product of univariate (one variate) density function. This is a function of $y$ with a fixed $\theta$. But the Likelihood function is the other way round, a function of $\theta$ with a fixed $x$. We want to find the model parameters that will maximise the likelihood function over the parameter space, which is

$$\hat{\theta} = argmax\hat{L}_n(\theta; y)$$

Hence the mode of all the parameter values. The $argmax$ means the argument which should give the maximum value from specific function. A good condition that tests to see if a max estimate exists is for the function of the likelihood to be continuous over a parameter space that is compact (closed and bounded). Otherwise, the likelihood function could increase infinitely. For convenience we can log the likelihood so that it is easier to differentiate (we will explain why this is important later). This will not affect our result because natural logarithm is a monotonically increasing function. This means that the general positive or negative correlation of the graph stays the same as well as the turning points. This is important as we need to find the maximum value of the log likelihood $L_n$ by our parameters $\theta$ to find a maximum or minimum for which the necessary condition are:

$$\frac{\partial l}{\partial \theta_1} = 0, \frac{\partial l}{\partial \theta_2} = 0, ..., \frac{\partial l}{\partial \theta_k} = 0$$

This briefly explains the theory behind it. It is often hard to find and exact answer. In general, no closed form solution for the problem is known and MLE is done using a simulation. For executing out Bayesian method

using a Poisson prior we have done this computationally as it is very difficult otherwise.

# Chapter 2

# Derivations and Proofs

## 2.1 Frequentist Approach Derivations

### 2.1.1 Expectation

In order to calculate N (the population size) we need to have a formula for the expectation of $X_m$ our random variable. This will be the mean observation for the maximum value of the samples. We use the probability mass function to estimate this.

$$P(X_m = l) = \frac{\binom{l-1}{m-1}}{\binom{N}{m}}$$

To calculate the expected value for a discrete random variable $X_m$ we use the formula $E(X) = \sum_{allx} xp(x)$. We can swap out $X_m$ for $l$ since they are the same here. So we have,

$$E(X_m) = \sum_{l=m}^{N} lP(X_m = l)$$

It is important to note the summation starts with $m$ because of the condition

$l$ cannot be less than $m$. Then we can sub in the pmf to get,

$$E(X_m) = \sum_{l=m}^{N} l \frac{\binom{l-1}{m-1}}{\binom{N}{m}}$$

In order to derive this we can use a few identities.
The First Identity is from the binomial coefficient formula $\frac{l!}{m!(l-m)!}$.

$$\binom{l}{m} = \frac{l!}{m!(l-m)!} = \frac{l}{m}\binom{l-1}{m-1}$$

Since $l! = l(l-1)(l-2)... = l(l-1)!$ and similarly for $m$ we can take out the first $l$ and $m$ of the coefficient.
The Second Identity we will use is this,

$$\sum_{l=m}^{N} \binom{l}{m} = \sum_{l=m}^{N} (\binom{l+1}{m+1} - \binom{l}{m+1}) = \binom{N+1}{m+1}$$

We can use these identities to get a formula for the expectation. First we take the formula for $E(X_m)$ and multiple it by $m/m = 1$ so that we can re order the formula to get:

$$E(X_m) = \sum_{l=m}^{N} l \frac{\binom{l-1}{m-1}}{\binom{N}{m}} \frac{m}{m} = \sum_{l=m}^{N} \frac{m}{\binom{N}{m}} \frac{l}{m} \binom{l-1}{m-1}$$

As you can see this makes it easy for us to substitute the first identity into the equation to get:

$$= \sum_{l=m}^{N} \frac{m}{\binom{N}{m}} \binom{l}{m}$$

Then we can sub in the second identity to get:

$$= \binom{N+1}{m+1} \frac{m}{\binom{N}{m}}$$

By expanding the binomial coefficient we get

$$= \frac{m(N+1)!}{[(N+1)-(m+1)]!(m+1)!}\frac{(N-m)!m!}{N!} = \frac{m(N+1)!}{(N-m)!(m+1)!}\frac{(N-m)!m!}{N!}$$

We can cancel out $(N-m)!$ to get:

$$= \frac{m(N+1)!}{(m+1)!}\frac{m!}{N!}$$

We can also cancel $\frac{m!}{(m+1)!} = \frac{1}{m+1}$ and $\frac{(N+1)!}{N!} = N+1$to get:

$$= \frac{m(N+1)}{m+1}$$

Thus our expectation is

$$E(X_m) = \frac{m(N+1)}{(m+1)}$$

### 2.1.2 Point Estimate

Our goal is to find an equation and thus a value for N. We can get a point estimate N by re arranging the equation for expectation.

$$N = \frac{m+1}{m}E(X_m) - 1$$

Make N be point estimate Y. This is because as we do not know $E(X_m)$, we can introduce Y as a random variable which is dependent on $X_m$. It has an unbiased estimator for N so E(Y)=N.

$$Y = \frac{m+1}{m}X_m - 1$$

### 2.1.3 Variance

We know the equation for variance in terms of expectation.

$$Var(X) = E(X^2) - E(X)^2$$

We can evaluate the factorial moment $E(X^2)$ by

$$E(X^2) = E(X_m(X_m + 1)) - E(X_m)$$

This works out as $E(X_m(X_m + 1)) - E(X_m)$ would expand to $E(X_m^2) + E(X_m) - E(X_m)$ and since $E(X_m) - E(X_m) = 0$ it is still equal to $E(X_m^2)$. We can evaluate the factorial moment as such:

$$E(X_m(X_m + 1)) = \sum_{l=m}^{N} l(l + 1)P(X_m = l)$$

We can sub in the formula for pmf,

$$= \sum_{l=m}^{N} l(l + 1)\frac{\binom{l-1}{m-1}}{\binom{N}{m}}$$

In order for more clarity let us isolate the part $l(l + 1)\binom{l-1}{m-1}$ of the above equation to make it into $\binom{l+1}{m+1}$.First we divide by $m(m + 1)$.

$$\frac{l(l + 1)}{m(m + 1)}\binom{l - 1}{m - 1}$$

By using the identity for binomial coefficient $\binom{l}{m} = \frac{l!}{m!(l-m)!}$ we expand the equation to

$$\frac{l(l + 1)}{m(m + 1)}\frac{(l - 1)!}{(m - 1)!((l - 1) - (m - 1))!}$$

We can cancel out the part in the denominator to

$$\frac{l(l+1)}{m(m+1)}\frac{(l-1)!}{(m-1)!(l-m)!}$$

The numerator is $l(l+1)(l-1)! = (l+1)!$ because $(l+1)l(l-1)(l-2)... = (l+1)!$. Similarly the denominator we have $m(m+1)(m-1)! = (m+1)!$ to get:

$$\frac{(l+1)!}{(m+1)!(l-m)!} = \binom{l+1}{m+1}$$

Thus we have the identity

$$\frac{l(l+1)}{m(m+1)}\binom{l-1}{m-1} = \binom{l+1}{m+1}$$

Going back to the equation$\sum_{l=m}^{N} l(l+1)\frac{\binom{l-1}{m-1}}{\binom{N}{m}}$. We can use the identity we just worked out on the factorial moment expectation:

$$E(X_m(X_m+1)) = \frac{1}{\binom{N}{m}}\sum_{l=m}^{N}\binom{l+1}{m+1}m(m+1)$$

$$= \frac{m(m+1)}{\binom{N}{m}}\sum_{l=m}^{N}\binom{l+1}{m+1}$$

Remembering that $\sum_{l=m}^{N}\binom{l}{m} = \binom{N+1}{m+1}$ we can sub that into the equation

$$= \frac{m(m+1)}{\binom{N}{m}}\binom{N+2}{m+2}$$

By using the identity for binomial coefficient expansion we can expand $\binom{N+1}{m+1}$

and $\binom{N}{m}$ to :

$$= \frac{\frac{m(m+1)}{N!}}{m!(N-m)!} \frac{(N+2)!}{(m+2)!((N+2)-(m+2))!}$$

$$= \frac{m(m+1)(m!(N-m)!)}{N!} \frac{(N+2)!}{(m+2)!(N-m)!}$$

The $(N-m)!$ cancel out on the denominator and numerator. Hence of the left equations numerator there is only $m(m+1)m!$ left.By expanding this we get $(m+1)m! = (m+1)m(m-1)(m-2)... = (m+1)!$. This simplifies to $m(m+1)!$.

$$= \frac{m(m+1)!}{N!} \frac{(N+2)!}{(m+2)!} = \frac{(N+2)!}{N!} \frac{(m(m+1)!)}{(m+2)!}$$

By expanding the factorials $\frac{(N+2)!}{N!}$ to $\frac{(N+2)(N+1)N(N-1)...}{N(N-1)...}$ we can cancel the top and bottom to get $(N+2)(N+1)$. In a similar fashion the part $\frac{(m(m+1)!)}{(m+2)!} = \frac{m(m+1)m(m-1)...}{(m+2)(m+1)m(m-1)...}$ will cancel out to be $\frac{m}{(m+2)}$. So we get:

$$E(X_m(X_m+1)) = \frac{m(N+2)(N+1)}{(m+2)}$$

Going back to the Variance equation in terms of expectation we have.

$$Var(X_m) = (E(X_m(X_m+1)) - E(X_m) - E(X_m)^2$$

Using the expectation we calculated early on and the equation for $E(X_m(X_m+1))$ we have:

$$= \frac{m(N+2)(N+1)}{(m+2)} - \frac{m(N+1)}{(m+1)} - \frac{m^2(N+1)^2}{(m+1)^2}$$

We can join the last two equations by factorising out the expectation $\frac{m(N+1)}{(m+1)}$

to get

$$= \frac{m(N+2)(N+1)}{(m+2)} - \frac{m(N+1)}{(m+1)}(1 + \frac{m(N+1)}{(m+1)})$$

We can expand the numerator of the last equations $1 + \frac{m(N+1)}{(m+1)} = \frac{m+1}{m+1} + \frac{m(N+1)}{(m+1)} = \frac{m+1+m(N+1)}{(m+1)}$ so our equation is:

$$= \frac{m(N+2)(N+1)}{(m+2)} - \frac{m(N+1)}{(m+1)}(\frac{m+1+m(N+1)}{(m+1)})$$

We can factorise out the $m(N+1)$ to get:

$$= m(N+1)(\frac{(N+2)}{(m+2)} - (\frac{m+1+m(N+1)}{(m+1)^2}))$$

Factorise out m in the second fraction.

$$= m(N+1)(\frac{(N+2)}{(m+2)} - (\frac{m(N+2)+1}{(m+1)^2}))$$

Join the fractions together with common denominator.

$$= m(N+1)(\frac{(m+1)^2}{(m+1)^2}\frac{(N+2)}{(m+2)} - (\frac{(m(N+2)+1)(m+2)}{(m+1)^2(m+2)})$$

$$= m(N+1)(\frac{(m+1)^2(N+2) - m(N+2)(m+2) - (m+2)}{(m+1)^2(m+2)}$$

$$= m(N+1)(\frac{(N+2)((m+1)^2 - m(m+2)) - (m-2)}{(m+1)^2(m+2)})$$

$$= m(N+1)(\frac{(N+2)((m^2+2m+1) - (m^2+2m)) - (m-2)}{(m+1)^2(m+2)})$$

$$= m(N+1)(\frac{(N+2) - (m+2)}{(m+1)^2(m+2)}$$

Thus our equation for variance is

$$Var(X_m) = \frac{m(N+1)(N-m)}{(m+1)^2(m+2)}$$

### 2.1.4 Variance of Point Estimate Y

We take the variance of the point estimate equation calculated before $Y = \frac{m+1}{m}X_m - 1$.

$$Var(Y) = Var((\frac{m+1}{m})X_m - 1) = (\frac{m+1}{m})^2 Var(X_m)$$

We sub in the equation for Variance of $X_m$.

$$= (\frac{m+1}{m})^2 \frac{m(N+1)(N-m)}{(m+1)^2(m+2)}$$

Therefore, the variance is:

$$= \frac{(N+1)(N-m)}{m(m+2)}$$

## 2.2 Bayesian Approach Derivations

### 2.2.1 PMF

Using Bayes theorem $P(A|B) = \frac{P(B|A)P(A)}{P(B)}$ we can ascertain the equation for $P(N = n|X_m = l)$ the probability that $n$ is the true population value given the maximum value of the sample size is $l$.

$$P(N = n|X_m = l) = \frac{P(X_m = l|N = n)P(N = n)}{P(X_m = l)}$$

To get the likelihood equation we can use the law of total probability $P(A) = \sum_n P(A|B_n)P(B_n)$ on the denominator.

$$= \frac{P(X_m = l|N = n)P(N = n)}{\sum_{\infty}^{v=1} P(X_m = l|N = v)P(N = v)}$$

We know $P(X_m = l|N = n) = \frac{\binom{l-1}{m-1}}{\binom{N}{m}}$ if $m <= l <= n$ and 0 otherwise (N=n fixes N). We have an expression for the prior $P(N = n) = \frac{1}{\Omega}$ if $1 <= n <= \Omega$ and 0 otherwise. We use a uniform prior for this because we assume that in the context of collecting destroyed tank pieces with unique serial number that each tank piece is equally as likely to be picked as any other. So they all have the equal probability of $\frac{1}{\Omega}$ to be picked. Multiply the Likelihood and the Prior (as posterior is directly proportional to the product of likelihood and prior) to get:

$$= \frac{P(X_m = l|N = n)P(N = n)}{\sum_{\infty}^{v=1} P(X_m = l|N = v)P(N = v)} \frac{\frac{1}{\Omega}}{\frac{1}{\Omega}}$$

If we sub in the pmf

$$= \frac{\frac{\binom{l-1}{m-1}}{\binom{N}{m}}}{\sum_{v=l}^{\Omega} \frac{\binom{l-1}{m-1}}{\binom{v}{m}}} = \frac{\binom{n}{m}^{-1}}{\sum_{v=l}^{\Omega} \binom{v}{m}^{-1}}$$

Using the Identity $(\binom{n}{m})^{-1} = \frac{m}{m+1}(\binom{n-1}{m-1}^{-1} - \binom{n}{m-1}^{-1})$ we can evaluate the denominator of the posterior pmf

$$\sum_{v=l}^{\Omega} \binom{v}{m}^{-1} = \frac{m}{m-1} \sum_{v=l}^{\Omega} (\binom{v-1}{m-1}^{-1} - \binom{v}{m-1}^{-1})$$

Subbing in the limits

$$= \frac{m}{m-1}(\binom{l-1}{m-1}^{-1} - \binom{\Omega-1}{m-1}^{-1}) = \frac{m}{m-1}\binom{l-1}{m-1}^{-1}$$

The series converges, i.e. the limit $\Omega \to \infty$ exists as the last term vanishes in this limit. Hence the posterior pmf does not depend on the prior for sufficiently large values of $\Omega$. Following the equation from before we sub in the above identity.

$$= \frac{\binom{n}{m}^{-1}}{\frac{m}{m-1}\binom{l-1}{m-1}^{-1}}$$

If we just isolate the denominator and look at that $\frac{m-1}{m}\binom{l-1}{m-1}$ (switched to numerator). We can evaluate the binomial coefficient as such.

$$\frac{m-1}{m}\binom{l-1}{m-1} = \frac{m-1}{m}\frac{(l-1)!}{(m-1)!(l-m)!}$$

We can expand the denominator $(m-1)! = (m-1)(m-2)...$ and cancel out $m-1$ to get $(m-2)!$:

$$\frac{(l-1)!}{(m-2)!(l-m)!m}$$

If we multiple by $\frac{l}{l}$ since it equals 1 we can manipulate the fraction so $(l-1)! = l!$.

$$\frac{l}{l}\frac{(l-1)!}{(m-2)!(l-m)!m} = \frac{1}{l}\frac{l!}{(l-m)!(m-2)!m}$$

We can manipulate the denominator part $(m-2)!m$ by expanding the factorial and multiplying by $\frac{m-1}{m-1} = 1$ to get:

$$(m-2)!m\frac{m-1}{m-1} = \frac{m(m-1)[(m-2)(m-3)...]}{(m-1)} = \frac{m!}{m-1}$$

We will use this version of the pmf for the bayesian estimates.

$$= \frac{m-1}{l}\frac{l!}{(l-m)!m!} = \frac{m-1}{l}\frac{\binom{l}{m}}{\binom{n}{m}}$$

### 2.2.2 Expectation

We can find the expectation by computing the first moment of the pmf by multiplying n by the pmf.

$$E(N|X_m = l) = \sum_{n=l}^{\infty} nP(N = n|X_m = l) = \sum_{n=l}^{\infty} n\frac{m-1}{l}\frac{\binom{l}{m}}{\binom{n}{m}}$$

$$= \frac{m-1}{l}\binom{l}{m}\sum_{n=l}^{\infty}\frac{n}{\binom{n}{m}}$$

Let us isolate a part of the equation $\frac{n}{\binom{n}{m}}$. We can change this using the the binomial coefficient

$$\frac{n}{\binom{n}{m}} = \frac{n}{\frac{n!}{m!(n-m)!}} = \frac{nm!(n-m)!}{n!}$$

We can cancel out $\frac{n}{n!} = \frac{1}{(n-1)!}$ so we get

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

We can make $m! = m(m-1)!$ in order to transform the equation into a binomial coefficient

$$= \frac{m(m-1)!(n-m)!}{(n-1)!} = m\binom{n-1}{m-1}^{-1}$$

Putting this back into the $\frac{m-1}{l}\binom{l}{m}\sum_{n=l}^{\infty}\frac{n}{\binom{n}{m}}$ we get

$$= \frac{m-1}{l}\binom{l}{m}\sum_{n=l}^{\infty}\frac{m}{\binom{n-1}{m-1}}$$

We can use identity $\sum_{v=l}^{\Omega}\binom{v}{m}^{-1} = \frac{m}{m-1}\binom{l-1}{m-1}^{-1}$ to get

$$= \frac{m-1}{l}\binom{l}{m}m\frac{m-1}{m-2}\binom{l-2}{m-2}^{-1}$$

So the equation we get for expectation is

$$= \frac{m-1}{m-2}(l-1)$$

### 2.2.3 Variance

We know the equation for Variance in terms of Expectation.

$$Var(X_m) = E(N(N-1)|X_m = l) + E(N|X_m = l) - (E(N|X_m = l))^2$$

We can evaluate the first factorial moment

$$E(N(N-1)|X_m = l) = \sum_{n=l}^{\infty} n(n-1)\frac{m-1}{l}\frac{\binom{l}{m}}{\binom{n}{m}} = \frac{m-1}{l}\binom{l}{m}\sum_{n=l}^{\infty}\frac{n(n-1)}{\binom{n}{m}}$$

We can expand the binomial coefficient $\binom{n}{m}$

$$= \frac{m-1}{l}\binom{l}{m}\sum_{n=l}^{\infty}\frac{n(n-1)}{\frac{n!}{m!(n-m)!}}$$

We can isolate the last part $\frac{n(n-1)}{\frac{n!}{m!(n-m)!}}$ and calculate this.

$$\frac{n(n-1)m!(n-m)!}{n!}$$

By expanding the denominator $n! = n(n-1)(n-2)...$ we can cancel out

$n(n-1)$ on the top and the bottom.

$$= \frac{m!(n-m)!}{(n-2)!}$$

We can expand $m! = m(m-1)(m-2)!$ in order to put this in a binomial coefficient.

$$= \frac{m(m-1)(m-2)!(n-m)!}{(n-2)!} = \frac{m(m-1)}{\binom{n-2}{m-2}}$$

We can plug this back into the equation.

$$= \frac{m-1}{l}\binom{l}{m}\sum_{n=l}^{\infty}\frac{m(m-1)}{\binom{n-2}{m-2}}$$

We can isolate the denominator and use identity $\sum_{v=l}^{\Omega}\binom{v}{m}^{-1} = \frac{m}{m-1}\binom{l-1}{m-1}^{-1}$ to get

$$\sum_{n=l}^{\infty}\binom{n-2}{m-2}^{-1} = \frac{m-2}{m-3}\binom{l-3}{m-3}^{-1}$$

Which when you put it back into the equations becomes:

$$= \frac{m-1}{l}\binom{l}{m}m(m-1)\frac{m-2}{m-3}\binom{l-3}{m-3}^{-1}$$

Thus,

$$E(N(N-1)|X_m = l) = \frac{m-1}{m-3}(l-1)(l-2)$$

Going back to the variance equation in terms of expectation and factorial moments

$$Var(X_m) = E(N(N-1)|X_m = l) + E(N|X_m = l) - (E(N|X_m = l))^2$$

We can sub in the values

$$Var(X_m) = \frac{m-1}{m-3}(l-1)(l-2) + \frac{(m-1)(l-1)}{m-2} - \frac{(m-1)^2(l-1)^2}{(m-2)^2}$$

Make it one fraction by finding a common denominator and calculate the equations as such:

$$= \frac{(m-1)(l-1)(m-2)^2(l-2) + (m-1)(l-1)(m-2)(m-3) - (m-1)^2(m-3)(l-1)^2}{(m-3)(m-2)^2}$$

$$= \frac{(m-1)(l-1)[(m-2)^2(l-2) + (m-2)(m-3) - (m-1)(m-3)(l-1)]}{(m-3)(m-2)^2}$$

$$= \frac{(m-1)(l-1)[(lm^2 - 4lm + 4l - 2m^2 + 8m - 8) + (m^2 - 5m + 6)}{(m-3)(m-2)^2}$$

$$\frac{-(lm^2 - 4lm + 3l - m^2 + 4m - 3)]}{(m-3)(m-2)^2}$$

Thus the variance is

$$Var(X_m) = \frac{(m-1)(l-1)(l-m+1)}{(m-3)(m-2)^2}$$

# Chapter 3

# Simulation

We will use R to simulate different experiments to mimic the method of collecting tanks and finding the point estimate. We will fix population size N = 1000 and Sample Size m=20.

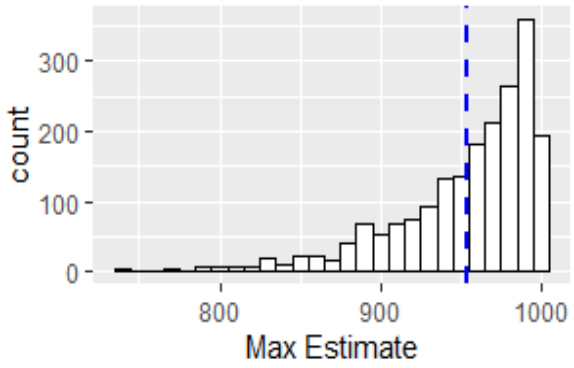
## 3.1 Biased Estimator

With the fixed variables we will generate 2000 samples (without replacement) and for each sample set we will find the sample max. We make a function to find the max of the sample of 20. The function inputs $x$ =the population and $m$ =samples size. It assigns the vector of 20 samples to a variable and return the max of it.

```
xsamples <- function(x,m){
xsamples <- sample(x,m)
return(max(xsamples)) }
```

We generate 2000 samples and store them in a column in a empty data frame. The results of this histogram that shows how many times a certain number is chosen. The blue line represents the average.

Figure 3.1: Graph of Maximum Sample Values




From this histogram we can obtain an estimator. Here we have that our outcomes are our sample maximums (highest serial number on captured tanks) that we plotted. We estimate the total number of tanks produced here.

Our point estimate can be represented as $n$. One estimator we have found is $n = maxx_i$. This is a biased estimator for our total population as the estimator is lower than the sample max obtained. The true population cannot be lower than the maximum serial number. This is because it is an underestimation and for it to be exactly the $n$ then it is a conservative estimate. We want an unbiased estimator( an accurate statistic that is neither a overestimate of underestimate).

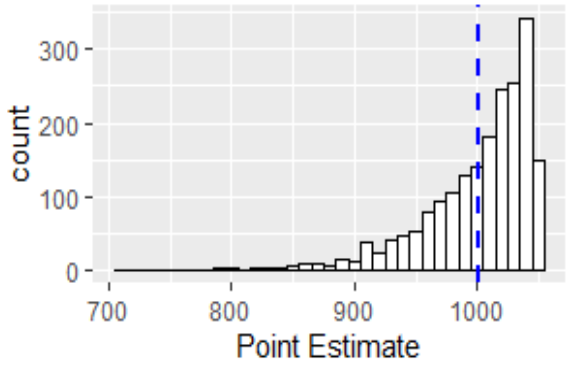
## 3.2 Frequentist Histogram

We can simulate this process of getting the point estimate by generating 2000 samples. We make a function where we input the sample max of each sample we generate and the fixed sample size.

```
N <- function(xmax, m){
  return(xmax*(1+(1/m))-1)}
```

We can display the point estimate on a histogram. The blue dotted line show the average value which is our point estimate.

Figure 3.2: Frequentist Approach Samples




We find that our estimator $\hat{n}$ is unbiased as the true values is near the mean=1000.139 and the standard deviation=46.08504. The true value of the total population is extremely close to the average of 1000. The standard deviation is moderately low which would imply our estimate is fairly accurate as the biggest and smallest it could be is 1046 and 954 respectively. This is still not a bad estimate.

Figure 4, the chart below tests different samples sizes to see how this affects the mean and variance. As we would expect with a higher sample size the mean becomes close to the true population size 1000 and the standard deviation (and hence the variance) decreases significantly as we increase the sample size (as we would expect). This would make it more consistent and an accurate estimator.

Figure 3.3: Chart of different Sample Sizes m for Frequentist Method

| Sample Size | Mean | Standard Deviation |
|---|---|---|
| 20 | 998.9885 | 46.11423 |
| 50 | 999.8169 | 19.14051 |
| 100 | 1000.256 | 9.204072 |
| 200 | 1000.03 | 4.43628 |

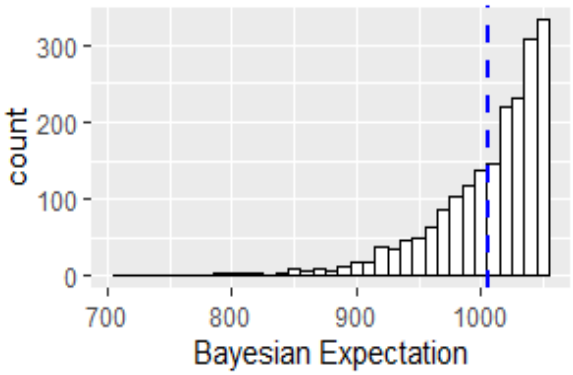
## 3.3 Bayesian Histogram

We can use the pmf to get the expectation for the Bayesian approach. Making our own R function for Bayesian expectations again we input the sample max and the sample size.

```
B_Exp <- function(xmax,m){
  res <- (xmax-1)*(m-1)/(m-2)
  return(res) }
```

We can use the Bayesian expectation of sample max as a point estimate. The blue dotted line show the average value which is our point estimate.

Figure 3.4: Bayesian Approach Samples




The mean of the histogram is 1005.381 and the standard deviation is 46.32887. Again we find the true value of the total population is extremely close to the average of 1005. The standard deviation is moderately low which would imply our estimate is fairly accurate as the biggest and smallest it could be is 1051 and 959 respectively. This is still not a bad estimate.
Similarly, to the Frequentist results. We see that as we increase the sample size the mean gets close to the true value of the population. As well as the standard deviation (variance) decreases with sample size.

Figure 3.5: Chart of different Sample Sizes m for Bayesian Method

| Sample Size | Mean | Standard Deviation |
|---|---|---|
| 20 | 1004.224 | 46.35822 |
| 50 | 1000.614 | 19.15614 |
| 100 | 1000.448 | 9.205931 |
| 200 | 1000.076 | 4.436503 |

## 3.4 Using a Different Prior

The prior probability is important to take into account as it affects the posterior. However we do not want it to largely affect the outcome. It is the probability distribution that would represent our belief before we take into account some new information or evidence. In our derivations so far we have assumed the distribution is Discrete Uniform. The Discrete Uniform Distribution is a symmetric probability distribution where the values are finite and have an equally likely probability of being observed. With a unique finite number of unknowns tanks (each with unique serial numbers) we can assume the prior is Uniform and can be sampled without replacement.

However we can also use a different prior such as negative binomial or Poisson. I will choose a Poisson distributions. The Poisson distribution is a discrete probability distribution which shows the frequency of times an event is likely to occur in a specific interval of time.

In the context of the German tank problem we can put it like this. Since the tanks we collect are those destroyed on the battlefield by the soldiers. Let's make an assumption that on the battlefield there were M tanks in total. The soldiers were able to knock out a fraction f of those tanks. Thus we would assume the wrecks on the battlefield expected number is $\mu = fM$.

One problem we see here is that it could be an underestimation or an overestimation. There could have been a significantly smaller number of tanks and the soldiers could have been very lucky in taking them all out. In contrast it would be possible that there were a large amount of tanks the soldiers were

not great at destroying them. This is were Poisson distribution would come in use.

Instead of making the parameter $\mu$ an integer number (amount of tanks taken out) we can relate it to a function of the probability of seeing a certain number of tanks being taken out. Hence, you get the pmf for Poisson distribution.

$$f(k;\mu) = Pr(X = k) = \frac{\mu^k \exp -\mu}{k!}$$

This means that as mu is the expected number, the probability of you observing an integer number k is represented by this formula.

## 3.5 Maximum Likelihood Estimator for Poisson Prior

To find a good estimate for what our $\mu$ should be we can take random guesses or consequently turn to the method of Maximum Likelihood Estimators. Say we want to find an estimate for $\mu$ in a Poisson distribution. We take a sample $Y_1, Y_2, ...Y_n$ $Poisson(\mu)$ from a Poisson distribution. We take the pmf of the distribution

$$f(y) = P(Y = y) = \frac{e^{-\mu}\mu^y}{y!}$$

We take the likelihood of this which becomes

$$L(f(y)) = \frac{e^{-n\mu}\mu^{\sum y_i}}{\prod_{i=1}^{n} y_i!}$$

Then we take the log of the likilihood

$$LogL(f(y)) = \frac{-n\mu + \sum y_i log(\mu)}{log(\prod_{i=1}^{n} y_i!)} = -n\mu + \sum y_i log(\mu) - log(\prod_{i=1}^{n} y_i!)$$

We can differentiate this and set this to 0

$$U = \frac{dl}{d\mu} = -n + \sum y_i(\frac{1}{\mu}) = 0$$

Since we want a formula for the parameter $\mu$ we can re arrange the above to get $\mu$:

$$\sum y_i(\frac{1}{\mu}) = n$$

$$\sum y_i = n\mu$$

Thus our parameter $\mu$ can be expressed as the average.

$$\mu = \frac{\sum y_i}{n}$$

## 3.6 Code

In this section we will explain the code used to generate this method. We use a sequence rangeP which ranges from all integer numbers from 0 to 2000 which will represent our x axis. We will use $l$ as our variable for max sample value estimate that we can change and experiment with. The sam is our sample size which we can also vary and change.

```
#parameters
library(ggplot2)
rangeP <- seq(0,2000)
l <- 1300 #max estimte l
sam <- 20 # sample m
```

Then we will work out the likelihood with our pmf and limits. We use the pmf worked out previously $P(X_m = l|N = n) = \frac{\binom{l-1}{m-1}}{\binom{N}{m}}$. This is the probability that the random variable $X_m$, that represents the maximum sample value will equal some value $l$. This is applied only in the condition that $l$ is between our sample size $m$ and our true population size $N$.

To get the values for this pmf we can run a while loop and use the sequence rangeP and apply the pmf conditions to it. We treat our rangeP as our random variable $l$. We make an empty vector "pmf" to fill with the values of our loop. We use q as the start point to run through the sequence (increases by 1 for every loop). The while loop condition states that while $q$ is smaller than or equal to the length of rangeP i.e.2000 the loop will continue to make sure all the values are calculated. The first $if$ statement is for probabilities that are equal to 0. That is if value $l$ is smaller than the sample size or bigger than the true population value then that value of $l$ will be given the pmf value NA which we later assign as 0. The else part will then apply the pmf to the remaining values.
Note that in the pmf equation we have $rangeP[q]$ rather than the population size N as we do not know the population size $N$. In the Bayesian method $N$ is not fixed it is a random variable.

```
#likilihood - pmf
q <- 1
pmf <- c(0) #empty vector to fill
while(q<=length(rangeP)){
if(l<sam|rangeP[q]<l){ #l<m & l>n
  pmf[q] <- NA #0 or NA
}else{
pmf[q] <- choose(l-1,sam-1)/choose(rangeP[q],sam) #pmf applied }
n <- n+1 }
pmf[is.na(pmf)] <- 0
```

We then work out the prior using the poisson distribution. Using the method of maximum likelihood estimation we can work out an estimator for $\mu$ that would be the average of the sequence rangeP. As well as testing out random other values for $\mu$ if we wanted. This *dpois* function is used to show the Poisson density in an R plot. It calculates the probability of a random variable that is in a certain range.

```
#prior
mu <- sum(rangeP)/length(rangeP)#log likelihood estimate
prior <-dpois(rangeP, lambda = mu)
```

Now that we have the likelihood and the prior we can compute a prior. Previously we have seen the posterior to be $P(A|B) = \frac{P(B|A)P(A)}{P(B)} = \frac{P(B|A)P(A)}{\sum P(B|A)P(A)}$. We can gauge the shape of what the distribution would be with just the numerator which would give what the posterior is proportional to $P(AB) \propto P(A)P(B|A)$. The denominator is used to normalise the curve. We then calculate 2 things, the unstandardised posterior (proportionality) and the standardised (normalised) posterior.

```
#dataframe
mydata <- cbind.data.frame(rangeP, pmf,prior,
pmf*prior,pmf*prior/sum(pmf*prior))
```

We can find an estimate with the unstandardised graph by fining the mode of the graph. This is the maximum value of the graph (the peak) we find this by using the Rcode:

```
mydata[mydata$`pmf * prior`== max(mydata$`pmf * prior`),]
```

We can find the max probability and its corresponding population number to be 1079.
Alternatively, when we normalise the curve, as expected get a similar distribution shape. To get the estimate for population size here we work out the area under the curve (hence the expectation). As our distribution is a discrete random one the formula $E(X) = \sum x * p(x)$. Hence we use the following R code:

```
sum(mydata$rangeP*mydata$`pmf * prior/sum(pmf * prior)`)
```

When we do this we get the result 1080.292.

## 3.7 Different Prior Method Errors

The one problem with this method is the prior distribution highly affects the posterior. This should not happen in an accurate representation of posterior distribution model. When we change the value of $\mu$ the expectation changes drastically.

## 3.8 Conclusion

We have many estimates deriving from different methods. Out of all the estimates the one from the Frequentist method has proven to be the most effective as it it closest to the true population. However, the other methods have proven to be just as good with fairly close estimates. We know from previous simulations that with an increase in sample size we have and increase in accuracy of results from both Frequentist and Bayesian approaches. This equally contributes to getting a good estimate.

# Chapter 4

# Bibliography

2020. 7.1 1 The German Tank Problem. [video] Available at: <https://www .youtube.com/watch?v=aHfbCL9gkW4t=8s> [Accessed 30 August 2021].

Birkett, A., 2020. Bayesian vs. Frequentist A/B Testing: What's the Difference?. [online] CXL. Available at: <https://cxl.com/blog/bayesian-frequentist-ab-testing/> [Accessed 30 August 2021].

Brooks-Bartlett, J., 2018. Probability concepts explained: Maximum likelihood estimation. [online] Medium. Available at: <https://towardsdatascience.com/probability-concepts-explained-maximum-likelihood-estimation-c7b4342fdbb1> [Accessed 30 August 2021].

Chatterjee, J., 2017. An efficient implementation of the German Tank Problem Statistical Algorithm in JAVA. International Journal of Scientific Engineering Research, [online] 8(6). Available at: <https://www.ijser.org/researchpaper/An-efficient-implementation-of-the-German-Tank-Problem-Statistical-Algorithm-in-JAVA.pdf> [Accessed 30 August 2021].

Glen, S., n.d. Estimator: Simple Definition and Examples. [online] Statis-

tics How To. Available at: <https://www.statisticshowto.com/estimator/> [Accessed 30 August 2021].

Hayes, A., 2021. Using the Variance Equation. [online] Investopedia. Available at: <https://www.investopedia.com/terms/v/variance.asp> [Accessed 30 August 2021].

Kenton, W., 2020. Probability Density Function (PDF) Definition. [online] Investopedia. Available at: <https://www.investopedia.com/terms/p/pdf.asp> [Accessed 30 August 2021].

En.wikipedia.org. 2021. Likelihood function - Wikipedia. [online] Available at: <https://en.wikipedia.org/wiki/Likelihood_function> [Accessed 30 August 2021].

Lima, F., 2019. Bayesian models in R | R-bloggers. [online] R-bloggers. Available at: <https://www.r-bloggers.com/2019/05/bayesian-models-in-r-2/>
Accessed 30 August 2021
.

En.wikipedia.org. 2021. Maximum a posteriori estimation - Wikipedia. [online] Available at: <https://en.wikipedia.org/wiki/Maximum_a_posteriori_estimation> [Accessed 30 August 2021].

En.wikipedia.org. 2021. Maximum likelihood estimation - Wikipedia. [online]
Available at: <https://en.wikipedia.org/wiki/Maximum_likelihood_estimation> [Accessed 30 August 2021].

En.wikipedia.org. 2021. Parametric family - Wikipedia. [online] Available

at: <https://en.wikipedia.org/wiki/Parametric_family> [Accessed 30 August 2021].

Corporate Finance Institute. 2021. Point Estimators. [online] Available at: <https://corporatefinanceinstitute.com/resources/knowledge/other/point-estimators/> [Accessed 30 August 2021].

GeeksforGeeks. 2021. Poisson Functions in R Programming - GeeksforGeeks. [online] Available at: <https://www.geeksforgeeks.org/poisson-functions-in-r-programming-2> [Accessed 30 August 2021].

profile, V., 2012. The German Tank Problem Revisited. [online] Cosmic-horizons.blogspot.com. Available at: <http://cosmic-horizons.blogspot.com/2012/09/the-german-tank-problem-revisited.html> [Accessed 30 August 2021].

Queen Mary University of London, Introduction to Statistics Notes, 2020
Simmering, J., 2014. Frequentist German Tank Problem | R-bloggers. [online] R-bloggers. Available at: <https://www.r-bloggers.com/2014/03/frequentist-german-tank-problem/> [Accessed 30 August 2021].

Simon, C., n.d. the German tank problem | The Simon Ensemble. [online] Simonensemble.github.io. Available at: <https://simonensemble.github.io/2019-11/german-tank-problem> [Accessed 30 August 2021].