Statisticsduring the period of the Second World War
-Lakshmi Chandrasekaran

It’s the year 1943 and the World War II is on in full swing.

The Allied forces managed to capture a few German tanks – a difficult task given their superior design. Now as you can imagine, there was considerable interest among the Allied camps to figure out how many such tankswere produced per monthby theGermans! That would help the Allied forces determine how to organize their resources.

Using traditional wartime techniques such as spying and decoding intercepted messages , the Allied intelligence community estimated that Germans produced anywhere between 1000 to 1400 tanks per month. However, much later,the Allied forces found that in reality the tank production estimate was unreliable and was nowhere close to that number. Was there another way they could“see into the shadows?” in order to determine an accurate number, asks Dr. Jonathan Farley, Associate Professor of Mathematics at Morgan State University in Baltimore, Maryland.

How statistics helped count German tanks without capturing them

The Allied forces used a mathematical approach to estimating the tank numbers. It started with the observation that each captured tank had a serial number. Assuming that a serial number indicates that the tanks were produced in a sequential manner (being a German production, they were correct in assuming that the enemy was methodical!) and every tank has an equal chance of being captured, how could one“guess” the total quantity of German tanks?

Since it is practically impossible to capture all the tanks, let us “assume” that the Allied forces captured a limited number of them. As a theoretical exercise let us assume,that the tanks were numbered as 1, 2, 3, 4….T, with T considered as the maximum serial number. Imagine that the Allied forces managed to capture fourtanks with serial numbers 1, 12, 15, 25.Then the problem was how to estimate the maximum number of a population (in the current example, it is the total number of tanks), given only the size of a sample?After several experiments a set of statisticians (a specialized branch of mathematics) came up with a relatively simple formula, in order to accurately estimate the population size:

> Population maximum= Sample maximum + (Sample maximum-Sample size)/Sample size

In our example,the sample maximum was 25 (assuming the highest serial number denotes the maximum number of tanks produced in that lot) and the sample size was 4 based on the number of captured tanks.Using the formula above, the statisticians’ estimation turned outthat the population maximum was 25+(25-4)/4, which approximately yielded the value 30. This calculation in Statistics falls under a topic known as “estimation theory”.

Statistics is a branch of mathematics used extensively in analysis, organization and classification of data. It has versatile real world applications in the field of medicine, financial markets, politics, war etc. In the post war period, when German war records were analyzed the actual number of tanks produced per month during wartime was found out to be 255. The estimation theory applied to the captured tanksproduced a number as 256, which was so close to the actual quantity, attesting to the power of Statistics.

But this wasn’t the only math that was applied during the war! Yet another branch of mathematics that was extensively used was probability, which is the science of predicting how likely an event would occur, since many a times in the real world it is hard to predict anything with certainty.One of the simplest examples of probability is that of a coin toss. A coin has 2 sides- a head and a tail. Once you flip a coin, then what is the probability of getting a head for instance? Since a coin toss could result in either a head or tail, the probability of getting either a head or tail is 50% or simply 0.5. Probability is a powerful concept very widely used in mathematics, statistics, artificial intelligence, medicine etc.

Probability saves planes from being shot down

Here is anotherwartime example thatillustrates the power of probability- During the 1940s, Abraham Wald was part of an elite group of mathematicians and statisticians known as Statistical Research Group (SRG) in the United States. Wald along with the other members of the SRG (affiliated with Columbia University, New York) was secretly housed in an apartment, where one of their objectives during the wartime was to figure outhow to make fighter planes immune to shooting by German forces. Was this not a job for an airplane engineer? Why ask a mathematician for the answer?!

The SRG looked into the bullet patterns found in the planes, which returned from the war field to see if they could be armored accordingly such that they were more immune to hits from the enemy. However, armoring them too much would make the planes heavier and would make them harder to fly. So what was the optimal way to minimize aircraft damage and loss?

“Wald's key observation was that the data on bullet holes represented only planes that had not been shot down. He figured that planes were getting shot at fairly uniformly, but the planes that were shot in the engine weren't returning to be seen,”says Dr. Rebecca Goldin, Director of STATS and Professor of Mathematical Sciences at George Mason University.

And so Wald proposed a counter-intuitive solution to the problem-instead of looking for bullet hole patterns in the plane, he suggested looking for parts of the plane, which were not bullet ridden.“To make this more useful, you can even break the airplane up into different parts (e.g., fuselage, engines, etc.) and determine how vulnerable planes are, given the damage to these areas,” says Farley.

Wald observed for instance that the aircraft fuselage suffered more bullet holes and not so much the engine, deducingthat the planes whose engines got a higher hit were actually not coming back. Now, why is this an important observation? As Jordan Ellenberg beautifully puts in his recently published book '*How not to be wrong'*- "If you go to the recovery room at the hospital, you'll see a lot more people with bullet holes in their legs than people with bullet holes in their chests. But that's not because people don't get shot in the chest; it's because the people who get shot in the chest don't recover," thus comparing the engine to be the "heart" of the plane. If you get shot in the heart you are less likely to survive than if you get shot in the leg!

This example underlies a very important concept that as humans we can be found guilty of and something often seen in business models – selection bias!Often times, when we try to analyze or understand a situation, we do not maintain an objective stance andour biases come into the picture, thus leadingus to prefer some outcomes over others. This bias, often leads to incorrect conclusions. In the case of the airplane problem, the selection bias was that others solely focused on parts of the plane hit by enemy bullets. However, Wald's astute observation was to pay attention to the parts not destroyed by bullets and to armor them so that they are better equipped to withstand enemy assaults! As the Harvard Business Review puts it, "The theoretically correct way to discover what makes a business successful is to look at both thriving and floundering companies."

Statistically speaking, what was the kind of problem that Wald was trying to solve? He was asking a simple question- how to determine the probability (chance) that the plane will survive certain number of hits, say four hits given that it has been hit thrice?Hypothetically speaking, if there were 10 planes and7 of them returned from the war field with say three bullet holes on them. Wald's clever observation was where exactly were these bullet holes on the plane - on the engine, fuel system, fuselage or other plane parts? He noticed that the planes that returned had lesser hits on the engine compared to other parts. So what about the other 3 planes that never came back? Could it be due to mechanical failure or enemy shooting? The most logical assumption would be since this was during war times, the 3 planes were shot down by the enemies and based on this Wald assumed that they were hit on the engines which is why they never returned! Wald used this data along with other mathematical and statistical theories to calculate the probability of an airplane survival.

And here's anotherpractical application of probability in something other than war- medicine.Let us assumea patient is alcoholic. Now what is the chance that they will be affected by liver disease? In order to find out, we would calculate a number of probabilities. First of all, let us say based on data from a clinic, we can estimate that the probabilityof patients coming to a clinic having liver disease is 'x'. The probability that patients coming to the clinic are alcoholic is 'y'.And let us say we know that among those patients with liver disease the probability that they are alcoholic is 'z'. Based on all these probabilities the equation (z*x)/y, popularly referred to as Bayes' Theorem, will tell us the chances (or probability) that if a patient is alcoholic, what are their chances of them having liver disease!

War and Mathematics

But these are examples for WWII. What about the impact of mathematics and statistics in contemporary combats? Farley's own research involves using concepts from Lattice theory, a sub-discipline of algebra. His study quantifies the probability that a given terrorist network has been disabled.Farley's doctoral student Zeinab Bandpey, at Morgan State University is trying to "see into the shadows" as Wald did.Using mathematical techniques, Bandpey is trying to answer the question,"How can you determine the "optimal" structure of a terrorist cell, when we can't really observe the cells themselves?" says Farley, whose work on counter terrorism has been used by Ministry of National Security of Jamaica and cited by researchers in China.

Another renowned wartime scientist from Britain, Alan Turing, made great advances in the field of number theory to break codes, which had a significant impact on World War II. "Keeping financial information and passwords safe involves encryption, which has seen huge developments thanks to the field of number theory in mathematics," says Goldin.

Using statistical techniquesserved the intelligence forces well when other traditional methods failed miserably and over-estimated the monthly tank production. Similarly Wald's incisive observation, thanks in part to his mathematically trained brain gave rise to some clever solutions to the airplane survivability problem. These examples highlight the impact of "theoretical"statistics and their unexpected real world applications.

You would think a war could only be won by a show of might. But it turns out there is a lot of mind and statistics involved as well!