Millersville University

The History and Application of Survival Models

A Senior Thesis Submitted to the Department of Mathematics & The University Honors Program

In Partial Fulfillment of the Requirements

For the University & Departmental Honors Baccalaureate

By

Kaleb M. Schmerge

Millersville, Pennsylvania

April 2025

This Senior Thesis was completed in the Department of Mathematics, defended before and approved by the following members of the Thesis Committee:

Dr. Richard Mikula, Ph.D (Thesis Advisor)

Professor of Mathematics

Millersville University

Dr. Baoling Ma, Ph.D

Associate Professor of Mathematics

Millersville University

Dr. Kevin Robinson, Ph.D

Associate Professor of Mathematics

Millersville University

Abstract

The purpose of this paper is to cover the history and the application of actuarial science, the survival models used, and the applications of these models. This paper will involve a brief history of the probability and statistics that lead up to the actuarial field, the history of actuarial science itself, and the models that are used. The models that are covered are the uniform distribution, the exponential distribution, the Gompertz-Makeham distribution, and the log-normal distribution. These are the core models that are used in actuarial science, and they help in understanding the life tables that are used heavily in this field. These life tables have led many developments in actuarial science and the models described can generate different life tables based on the application of them, as well as becoming the building blocks of many other medicinal and biological fields as the need for data on life-times of not just humans, but biological life as a whole, grows with time.

Acknowledgements

I am incredibly thankful for Dr. Richard Mikula, Dr. Baoling Ma, and Dr. Kevin Robinson for their support through this process. They were very informative and supportive during this project, and I would not have been able to complete this project without their guidance. They have helped me at every step of the way in the completion of this thesis. With all of their feedback and help with this project, they have helped create the best version of this thesis. My thesis advisor, Dr. Mikula, has been incredibly supportive during this journey. He has given me a wealth of information, not just for my thesis, but for my career as an actuary as well. During our weekly meetings, we would sometimes discuss the actuary exams, as well as the information for my thesis, and thanks to his help, I was able to pass my first actuary exam. I hope to continue to use the help he has given me to pass my other actuary exams and to further my career as an actuary.

I also want to thank my parents for their support. They have always supported me throughout my college journey, and they have continued to support me as I start my job search. Despite all of the setbacks and rough periods that I experienced while in college, my parents were always there to offer their love and support. Thanks to my incredible parents and my thesis committee, I was able to complete this project to the best of my ability.

Introduction

History of Probability

Actuarial science, while not everyone may know what this topic is, has been on quite the rise in recent years. With the advancements in technology, leaning into the information age, and the constant communication, people have pushed actuarial science into one of the core pillars of mathematics. Despite how recent developments in the actuarial field have been made, the foundation of this field has been set for a very long time. Actuarial science has its roots in statistics and probability, which has a long and well documented history, so in order to understand the history of actuarial science, some areas of the history of probability will need to be discussed.

Probability has been utilized for centuries, with the modern notion of probability developing during the 1600s in the context of gambling problems. This game of chance was documented in many books published by mathematical geniuses of this time, such as Girolamo Cardona, Blaise Pascal, Daniel Bernoulli, and Pierre de Fermat. These men were fascinated with these games of chance, and they were in awe of trying to find ways to optimize betting with dice rolls. An example of these problems is as follows:

A gambler has undertaken to throw a six with a die in eight throws. Suppose she has made three throws without success. What proportion of the stake should she keep if she is to give up her fourth throw?

This example was a "gambler's ruin" problem, written by Christianus Huygens, another mathematician of the period who was also interested in these kinds of gambling problems

(Haberman 2). Problems like these were frequently found in other mathematical books and were later compiled and repurposed into the famous *Doctrine of Chances*, published by Abraham de Moivre. For the purposes of actuarial science, this is where there is a bit of a split between the history of actuarial science and statistics. As probability moved more into these games and principles, actuarial science moved more into the financial realm that these games of chance had to offer.

History of Actuarial Science

Before diving directly into when actuarial science was created, one topic will need to be discussed first, as it was the basis of the creation for actuarial science, that is, insurance. The first recorded use of insurance occurred in the 14th century in Genoa. At the time, this was used to avoid charges of usury, but either way this is the first known record of insurance. Moving forward into 1583, this is the first known case of life insurance being taken out. This took place in London where William Gibbons took out insurance for a twelve month time period. Unfortunately, Gibbons passed nearly a year later, but the insurance company made the claim that, if every month was twenty-eight days long, Gibbons would have lived the full term of the life insurance plan. This led to the company claiming that they did not need to pay the amount. Due to this discrepancy with no clear outcome, many major cities in Europe began to publish their own mortality tables every week during the 16th and 17th centuries. By the beginning of the 17th century, it was much more common to see insured ships for cargo and for larger businesses in Europe. This introduction and increase in insurance use led to the creation and need of actuarial science.

With insurance policies in their infancy, there was an increasing interest in demographic analysis. The pioneer of this field, John Graunt, created the foundation of, as an article of the Actuary Magazine states, "...the basis for analyzing longevity and death in a population group - or cohort - of people of the same age," (Gupta 14). This work dealing with this field led to the creation of the first life table, which will be covered in much more detail later. With this work, Edmond Halley, who was another major contributor to actuarial science, used these life tables that were becoming more popularized, and was able to find a way to calculate premiums for insurance given the age of the person who is paying for the insurance. This is something that has held to this day, with many actuaries still calculating and using it all over the world.

With the foundational work that has been laid out by both Halley and Graunt, the properties and field of actuarial science were becoming completely fleshed out. Some of the major turning points for this field were in the 1800s, with the creation of the post of actuary was enacted in 1821 to the National Debt Office, and the appointment of the first actuarial position in the government, who was the president of the Institute of Actuaries, and became the first actuary to the National Debt Office starting in 1822.

Thanks to the work that was started many years ago, actuarial science has flourished as a modern day profession. The duties of actuarial science have increased from just pension plans and premiums, which are still staples of the field today, to risk assessments using probability models, and so much more. Actuarial science has moved beyond just insurance work too. Many actuaries today fill much needed roles in government, investors for companies to help fund research projects, and actuaries have become more widely used in the general public.

Overall, actuarial science has a rich history that spans over hundreds of years. With the rise of insurance and expansion of trade, more insurance policies led to the expansion of insurance and the rise of actuarial science. This field has continued to move forward as insurance and risk management are still on the rise. Most developed countries rely heavily on actuaries and financial managers to function, and this is a trend that will continue for the foreseeable future.

Notation Differences

Probability Notation

Between probability, statistics, and actuarial science, they contain many similarities, as they all have similar histories and are used in similar areas in practice. However, the major difference between the fields is their notation. The actuarial notation that was created is meant to be a "short-hand" notation compared to the notation that is used in probability and statistics. This allows for actuaries to more efficiently work with interest rates and life tables. Examples of these differences will follow once the variables are defined. With many applications of the actuarial notation being used for life tables, that will be the framing that will be used when discussing the definitions of these formulas and the notation differences below.

Consider (x), as this will denote a life whose age is x. In this frame, the death of (x) can occur at any age greater than x, as well as future lifetimes models of (x), and $x + T_x$, where T_x is a random variable, can be developed through this framing.

F_x will represent the cumulative distribution function of T_x, and F_x will be known as the lifetime distribution from age x. Generally,

$$F_x(t) = P(T_x \leq t).$$

This is defined as the probability that someone who is age x will continue to live for at most *t* more time. This is the statistics notation with the actuarial framing. The actuarial notation is as follows:

$$ {}_{t}q_{x} = P(T_{x} \le t) = F_{x}(t). $$

Continuing with this framing, the survival function can also be defined as,

$$ S_{x}(t) = 1 - F_{x}(t). $$

This is the probability that someone who is age x will live for at least *t* more time (typically years). The equivalent actuarial notation is, ${}_{t}p_{x}$.

When considering t = 1, both ${}_{t}q_{x}$ and ${}_{t}p_{x}$ can be written as q_{x} and p_{x}, and when considering t = 1, q_{x} is typically referred to as the mortality rate at age x.

Going back to the cumulative distribution function, its derivative, the probability density function ($f_{x}(t)$), also has a different notation, which also relates to the survival function defined above. It can be denoted as:

$$ f_{x}(t) = \frac{d}{dt}F_{x}(t) = -\frac{d}{dt}S_{x}(t). $$

The terms that will now need to be shown relate are T_{x}, T_{0}, S_{x}, and S_{0}. We know that

$$ P(T_{x} \le t) = P(\{T_{0} \le x + t)\}|\{T_{0} > x\}) = \frac{P(x<T_{0}\le t+x)}{P(T_{0}>x)}. $$

This implies

$$ F_{x}(t) = \frac{F_{0}(x+t)-F_{0}(x)}{S_{0}(x)}, $$

as well as

$$ S_{x}(t) = \frac{S_{0}(x+t)}{S_{0}(t)} \Rightarrow S_{0}(x + t) = S_{0}(x)S_{x}(t). $$

This formula can be generalized using x, t, u all greater than or equal to zero as follows:

$S_x(t + u) = S_x(t)S_{x+t}(u)$, or in the actuarial notation:

$${}_{t+u}p_x = {}_tp_x \cdot {}_up_{x+t}.$$

This generalization is a helpful tool when determining expected values and moments for the survival function, specifically for lifetime distributions. These expectations will need some parameters that must be satisfied before continuing with the expectation of the survival function.

First, $S_x(0) = 1$. This satisfies the condition that the probability that a life at the current age, x, will live at least zero more years, which is one.

The second condition is $lim_{t\to\infty} S_x(t) = 0$. This is more of a common knowledge condition, as all lives will eventually end at some point, so the probability that someone of age x will never die is zero.

The third condition that for any x ≥ 0, $S_x(t)$ must be a decreasing function of *t*. This means that the probability of surviving one more year continues to get lower as the person of age x continues to become older.

Expectation Notation

The next kinds of notations that will be discussed involve expectation. They will relate the expected values, variances, and some of the formulas that are equivalent to one another.

$$E[T_x] = \overset{\circ}{e}_x$$

$$\overset{\circ}{e}_x = \int_0^\infty S_x(t)dt$$

where $\overset{\circ}{e}_x$ is known as the complete expectation of life at birth. The second moment is as follows:

$$E[T_x^2] = 2\int_0^\infty tS_x(t)dt.$$

Given that $Var[T_x] = E[T_x^2] - (E[T_x])^2$:

$$Var[T_x] = 2\int_0^\infty tS_X(t)dt - (\int_0^\infty S_x(t)dt)^2.$$

The above results are all incredibly useful when dealing with the different models in actuarial science. These are the basic building blocks that have been developed for many years, and with these notations and definitions, this allows for the models to take shape and allows for them to be used in their own respective fields.

Actuarial Models

The notation that was derived and defined above is applied to various different models within Actuarial Science. All of these models involve different parameters that can be applied to varying populations and scales. While many of the models involve insurance premiums, pension plans, death benefits, and the lives concerned range from human lives to electronic lifetimes. Each individual model has their own sets of parameters, as well as historical backing for why these models work.

Uniform Distribution

The uniform distribution can be defined as a continuous probability distribution where each equidistant interval has equal probabilities, and there are 2 parameters: a and b. The probability density function and the moments of this distribution are as follows:

$$f(x) = \frac{1}{b-a} \text{ for } a \leq x \leq b,$$

$$E[X] = \frac{b+a}{2},$$

$$Var[X] = \frac{(b-a)^2}{12},$$

which is defined on the interval of negative infinity to infinity. Due to the equal probabilities for every equidistant interval, this is not an ideal distribution to model lifetime data of individual people. This data set is used more often for short ranges of time or ages.

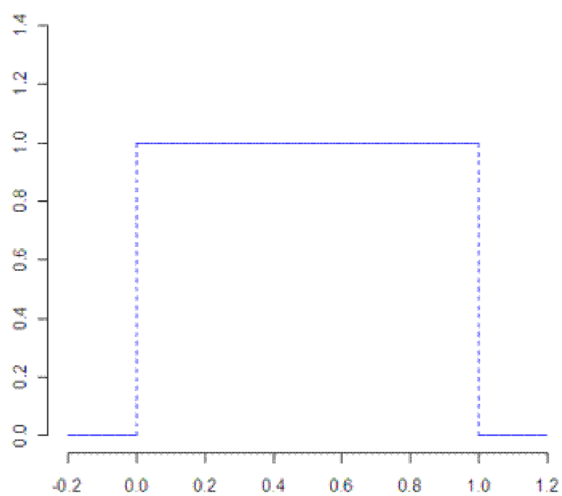

Figure 1. Uniform Distribution (*Essential Probability*)

The most notable use of this model was in 1724, when Abraham de Moivre proposed this method first to be used to model human survival functions. This has led to the uniform distribution to also be referred to as "de Moivre's law" (Cunningham 68).

Exponential Distribution

The exponential distribution can be defined as a continuous probability distribution with one parameter. The probability density function and the moments of this distribution are as follows:

$$f(x) = \frac{1}{\theta}e^{-\frac{x}{\theta}} \text{ for } x > 0,$$

$$E[X] = \theta,$$

$$Var[X] = \theta^2,$$

which is defined on the interval from zero to infinity. While this distribution can be used to model human lifetimes, it cannot be used over large ranges of time. The exponential distribution is similar to the uniform distribution as it is only used for short periods of time when modeling human life times. However, the exponential distribution can be used with the hazard function, which will be more clearly defined later.

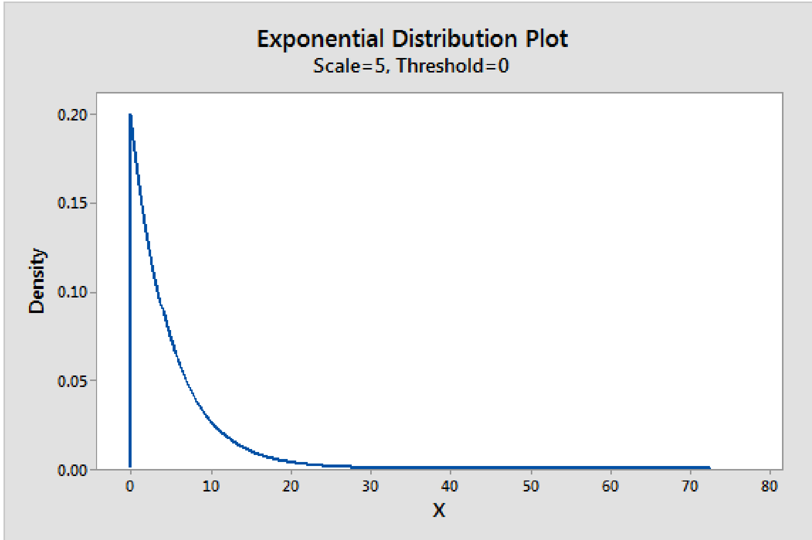

Figure 2. Exponential Distribution (*Frost*)

When involved with the hazard function, which will be discussed more later, the exponential distribution can be referred to as the constant force distribution. This distribution has the property of a constant hazard rate, and this is very useful when modeling inanimate objects, like electronics or machine parts. This distribution is incredibly useful in engineering for these reasons.

Gompertz-Makeham Distribution

The Gompertz-Makeham distribution can be defined as a continuous probability distribution that is specifically used to approximate lifetime lengths using demographic data. The survival function is as follows:

$$F_{\Theta}(s) = e^{-\alpha s - \beta(\frac{e^{\Upsilon s} - 1}{\Upsilon})},$$

where s is the lifetime, so $s \geq 0$, and α, β, and Υ are all nonnegative parameters.

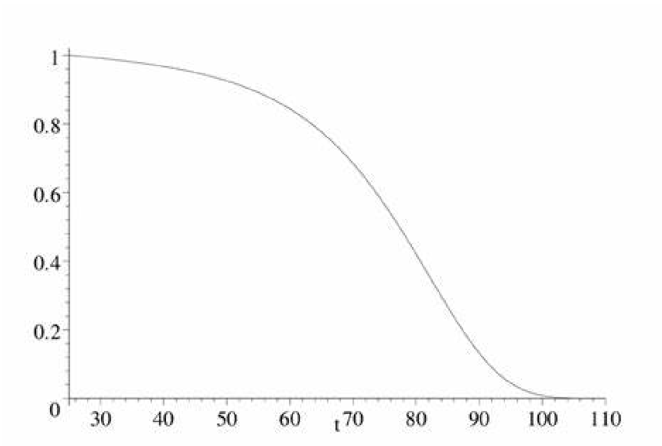

Figure 3. Gompertz-Makeham Distribution (*Menoncin*)

One of the main purposes for the Gompertz-Makeham distribution is used for insurance companies. This distribution gives very accurate probabilities on human lifetimes, and this allows for insurance companies to better estimate the costs for insurance plans and other fees that involve the length of human lifetimes (Norström 5). Insurance is not the only purpose for this model though. Other different professions may need to know human life expectancy given certain conditions for different treatments. The main benefit and use of this distribution is it creates a life-table, and this is revolutionary in all of the fields discussed above, as it allows for approximations to be easily read and makes determining conditional lifetime probabilities very intuitive to solve for.

The purpose for the Gompertz-Makeham distribution does not have to just involve human life times, as it can also be used to generate data on the lifetime of plants, crops, and could even be used on animals with a fairly long life expectancy. These applications make this an appealing distribution to use in biology, which has helped out agriculture in a multitude of ways. Knowing the probabilities of lifetimes and deaths of certain plants has allowed many farms to optimize their output, leading to an increase in food and efficiency for farmers globally. This model can be applied and changed for certain countries and average weather conditions, so this model can be repurposed around the world, and has been for many years in many fields, not just agriculture and biology.

Log-Normal Distribution

The log-normal distribution (or Galton distribution) can be defined as a continuous probability distribution where the logarithm of that distribution is distributed normally. The probability density function and the moments of this distribution are as follows:

$$f(x) = \frac{1}{\sqrt{x\sigma}} e^{\frac{-(lnx-\mu)^2}{2\sigma^2}} \text{ for } x > 0,$$

$$E[X] = e^{\mu+\frac{1}{2}\sigma^2},$$

$$Var[X] = e^{2\mu+\sigma^2}(e^{\sigma^2} - 1),$$

where μ is the mean of the natural logarithm, σ is the standard deviation of the natural logarithm, and this distribution is defined on the interval from 0 to infinity. This distribution is most commonly used in biology, finance, and environmental science. More specifically in actuarial science, the log-normal distribution can model various naturally occurring actions like income distribution, which can be used when calculating insurance plans based on income (Pavlovic 2).

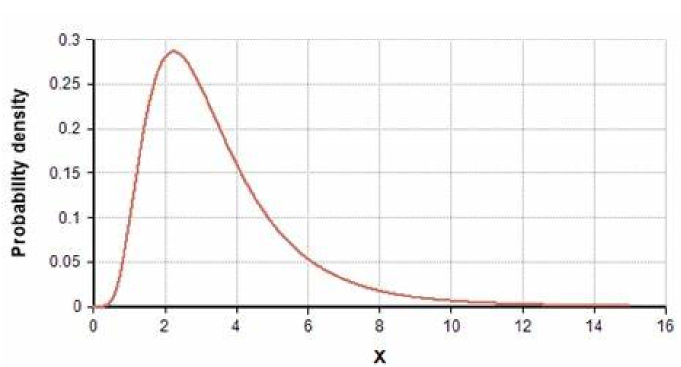

Figure 4. Log-Normal Distribution (*Lumina*)

Force of Mortality

Before transitioning over to the applications of the above models, one more important topic must be discussed, which is the force of mortality. The force of mortality, which is also referred to as the hazard rate or the failure rate, is defined as the instantaneous effect of mortality at a certain age. More generally, in the terms of the survival function, it is notated as $\mu(x)$. The notation for the force of mortality is as follows:

Recall that $F(x) = 1 - S(x)$. So,

$$\mu(x) = \frac{-s'(x)}{s(x)} \Leftrightarrow S(x) = e^{-\int_{x_0}^{x} \mu(t)dt}.$$

In general, if the force of mortality is large, that means that there is a larger probability that someone at age x will die in an upcoming short period of time. Another feature of this relationship is the force of mortality and the survival function each determine the other, however, the force of mortality tells us an instantaneous rate of death per unit of time, whereas the survival function focuses on not dying and shows the probability of living for at least a certain amount of time. These properties are incredibly useful information to know about a population, and the force of mortality can be applied to the models described in previous sections. For example, in the exponential distribution, the force of mortality is constant. This means that for a life of age x, the probability that a life will die in a short time is the same probability that a life of age x + c, where c is some arbitrary number, will die in a short time. While this is not useful for graphing human lives, as the force of mortality rate is not constant for humans or live beings, this is useful in electronic equipment. The equipment will work, with the same probability of working each time it is used, until one day the equipment does not work anymore. This concept is useful in the different models that were discussed above, and all of these pieces work together to form what is known today as life tables.

Life Tables

Development of Life Tables

The life table has had years of research and development in order to be the most accurate way of modeling human lifetime. It can show all the probabilities of death at certain ages, and someone's expected lifespan at all different ages. With all of these years of research, life tables have become incredibly accurate, but that came with time and centuries of data to back up these probabilities.

The first documented interest of this probability of death wasn't until the 17th century, where the first prototype of a life table was constructed by John Graunt in his article called *Natural and Political Observations Made upon the Bills of Mortality*, in London (Coale and Demeny 4). As expected, this was a very rough idea about the actual probability of human life time and the rate of mortality, but this article at the very least sparked people's interest in the subject. This article also normalized the format that all life tables use now, with the columns of death survival rates. Once this article was published, it took nearly forty years before someone else published their own version of a life table. This new life table was published by Edmund Halley, and it was written specifically to cover the years of 1687-1691 in the city of Breslau, Poland. These advancements led to other scholars creating and refining life tables as the years went on. Continuing on until around 1750, there emerged three different types of life tables. These three types of life tables consisted of data supplied by the registration system, census data, and a combination of data from registrations systems and a census.

In the first type of life table, the information supplied by the registration system contained different documents about births and deaths. These would be published in hospital records and other things of that nature. Since the documents were published by hospitals, the data that was given was incredibly reliable when constructing these specific life tables. The hospitals would also provide the distribution of the ages of the people that had passed, and this again gave reliable information when constructing the tables to include all ages in this research. Unfortunately, there were some major drawbacks to this method. This model only included and worked well with a population that was stationary. It did not account for migration, and for sudden deaths of individuals, there was not always an accurate way of determining the age of someone who had passed suddenly. Another major issue was this model could only be used in urban areas, and they could not be generalized for other urban areas. It was more or less a case study of their nearby city and was not able to be used much outside of that. Due to these overwhelming drawbacks, this model was considered worthless in further construction of life tables.

The second class of life tables, being constructed from a census, were a little different compared to the first model. The census allowed for more data to be collected on a wider scale, and this allowed the model to be partially generalized to a wider population. The way that this was conducted was by considering two censuses that were not taken too far apart, and the population numbers could be compared in order to construct a life table. This method allowed for a larger generalization compared to the first model, and it even had some use in non-European countries. Unfortunately, this only proved to work well in theory. Comparing censuses between countries quickly became difficult, as different countries had different standards for accurately completing a census, so this led to a large disparity between many different countries, specifically in India and Egypt. Another area of concern took form when it was discovered that it could not take into consideration births that occurred in between the time of the censuses, so the second census had inflated numbers. This led to varying inaccuracies as deaths and births could not be easily distinguished. Thus, again, this second model of constructing life tables was considered worthless.

This third model, which was a combination of the previous two models, was first created for the population of Sweden between 1755-1763. This type of model seemed to be more open and accessible, as there was large amounts of information that was available to construct these types of life tables. Once the first model of this type was constructed in Sweden, more researchers pulled together the data from their own populations and would create life tables of their surrounding areas, therefore, popularizing the use of this third type. There was a large interest that developed during the beginning of the 19th century, as more population statistics were created and censusses became more popular and accurate. Continuing on into the mid-20th century, there was a great multitude of life tables that were created, and due to the number and the interest, the tables were brought together at the *Office of Population Research* of Princeton University as a way to submit this type of life table as the standard for all future life tables.

When these models were overviewed at Princeton, there were many that did not fit the criteria that was needed in order to continue on with the research. Some of the constructed life tables did not differentiate between men and women, so these life tables were discarded. Other life tables overlapped and had been constructed for the same years, and these life tables were discarded. One of the main issues that appeared when examining each of the life tables was the overrepresentation of European countries and the underrepresentation of other countries. This led to heavy bias in life time data in these types of life tables. The researchers at Princeton attempted to resolve part of this issue though, as they decided it was best to use data that covered countries as a whole, so nearly every regional life table was discarded. This cut back on the number of European life tables that were in consideration for the final approval, but there was still a large bias of European life tables. Once every life table was examined, there were 326 models that remained.

From the remaining models, the researchers would then try to find some mathematical expression that could be used to explain the results from the life tables. There were many different approaches that were taken to find some correlation between the force of mortality, and how some expression could relate these results to every person in a single country. With some trial and error, the researchers were able to combine some models together in order to properly create the first graph of a life table. These models were the least squares models, the correlation values between the life tables, and the force of mortality of the Gompertz-Makeham distribution. The least squares model was used for a baseline of how the lifetimes should look on a graph, with the correlation checking to make sure none of the models fell too far outside of what was deemed a “good" range for the data given from the models. Finally, the Gompertz-Makeham distribution filled in the gaps of the graph. These three models led to the creation of the first fully completed life table. This was a major leap into the standardization of how life tables would be constructed moving forward. From this discovery, many other researchers would begin to develop life tables of their own nations, with an example of a life table graph as follows:

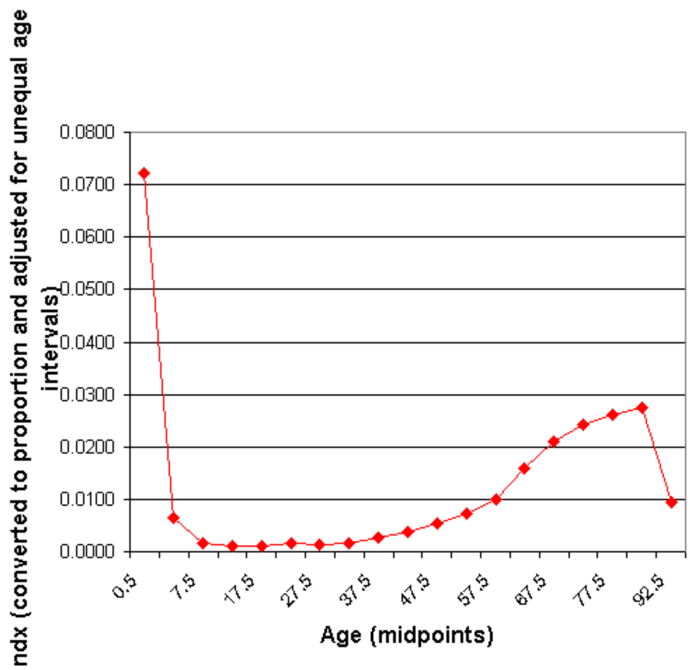

Figure 5. Life Table Distribution (*Measure Evaluation*)

This graph is the age distribution of deaths of Costa Rican males in the 1960s. Reading this graph, the probability of death is on the y-axis, with the ages on the x-axis. The dots in the graph are the different points where the probability of death is evaluated at given ages. The left side of the graphs show that a higher proportion of lives end in infancy, which is an unfortunate reality, but thankfully with modern medicine, this end of the graph has heavily decreased. The graph then levels off for many years as people become teenagers and enter adulthood. This graph shows that for males in the 1960s in Costa Rica, the probability of death is very low from around 10 years of age until 40 years of age. From this point, as people get older, their quality of life decreases and as their bodies age, the probability of death increases more and more the older people become, at least in this regional graph. The part to notice though from the graph is the ages between the ages of 80 to 95. The reason that this part of the graph turns downwards is actually because of how few people lived to be that age for this region. Because of that, there is a very small sample of people to analyze, and this leads to some skewed data towards the end of the graph.

Graphs of the life tables can be curated to what will best fit the population that is being examined, so many of these graphs will have varying graphs from nation to nation. The graph above is just one of many examples that have been created since the first creation of the life table graphs. Moving on from the graphs, the next step in understanding the information embedded in a life table and its graph is to understand the notation and the format given to the actual tables.

Life Table Notation

As described above, a life table is a reflection of the probability distribution of a life model. Typically, a life table will have a range starting at zero and continuing on until the corresponding survival function sees fit. This ranges from distribution to distribution, so there are differences in the life tables that can be constructed. The main examples that will be shown are models that deal with human lives, which means that these life tables are created with the Gompertz-Makeham distribution. Before looking at the table, some notation and definitions will need to be defined.

To first construct a life table, researchers will need the data that is given by ${}_tp_x$, which to recall, is the probability that a life at age x will live t more years. These values allow researchers to construct a life table for every age that they see fit, whether it is in intervals of one year, five years, or whatever is relevant to the researchers.

To further explain the notation involved in the construction of life tables, an example will be displayed below.

AUSTRALIAN LIFE TABLES 2015-17: MALES

Age	l_x	d_x	p_x	q_x	μ_x	$\mathring{e}_x$	L_x	T_x
0	100,000	355	0.996453	0.003547	0.000000	80.76	99,680	8,075,879
1	99,645	28	0.999721	0.000279	0.000362	80.05	99,630	7,976,212
2	99,617	17	0.999832	0.000168	0.000210	79.07	99,609	7,876,581
3	99,601	14	0.999862	0.000138	0.000147	78.08	99,594	7,776,973
4	99,587	10	0.999900	0.000100	0.000118	77.09	99,582	7,677,379
5	99,577	9	0.999912	0.000088	0.000091	76.10	99,573	7,577,797
6	99,568	8	0.999917	0.000083	0.000085	75.11	99,564	7,478,225
7	99,560	8	0.999917	0.000083	0.000083	74.11	99,556	7,378,661

Figure 6. Life Table Example (*BASHing data*)

This is just a small piece of a life table, but this has all of the sufficient parts needed for a standard life table. The first column is self explanatory, it is the age in which all of the probabilities relate to in each row.

The second column, represented by l_x and is the first number in the column, is referred to as the radix of the table. Moreover, l_x represents the number of people that are alive at any age x. As shown above, as the age increases, the value of l_x decreases, which falls in line with the model. This relates closely to the next columns, denoted as d_x, which represents the number of deaths that occurred between one value of *x* and the next. Something to note as well, is that in this short span of ages zero to seven, the number of deaths decrease at each interval, which is a similar pattern that was presented in the graph that involved the Costa Rican males. In order to solve for d_x, a simple calculation is needed:

$$d_x = l_x - l_{x+1}.$$

If a researcher wanted to determine the amount of death between some *n* amount of years, that could also be determined with the following formula:

$${}_n d_x = l_x - l_{x+n}.$$

Recalling some terms from earlier, ${}_nq_x$ represents the probability that someone who is age *x* will continue to live for at most *n* more years, and ${}_np_x$ represents the probability that someone who is age *x* will live for at least *n* more years. Both of these probabilities are related to the number of deaths that occur between each year. These are:

$$d_x = l_x \cdot q_x,$$

$${}_nq_x = \frac{{}_nd_x}{l_x}, \text{ and}$$

$${}_np_x = \frac{l_{x+n}}{l_x}.$$

Again, these formulas help construct the columns on the life table given the population statistics for a specific region.

For example, consider the interval of lives aged one and two. This means that out of the 99,645 lives that lived to age one on this table, 99,617 are expected to live to age two, with seventeen deaths occurring between ages one and two. Knowing these key pieces of information, the following are easily solved:

$${}_1p_1 = \frac{l_2}{l_1} = \frac{99617}{99645} = 0.999719,$$

which means given a life aged one, the probability of living to age two is 99.97%, and

$${}_1q_1 = \frac{d_1}{l_1} = \frac{28}{99645} = 1 - {}_1p_1 = .000281,$$

which means that the probability of death between age one and age two is 0.0281%.

Continuing on, as shown by the table, the force of mortality, μ_x, is also calculated in these life tables. This is calculated by the number of people that live between year to year, so it is not a constant rate, which is expected when dealing with people. This is given by the following formula:

$$\mu_x = -\frac{\frac{d}{dx} l_x}{l_x} = -\frac{d}{dx} ln(l_x).$$

To consider the force of mortality between an interval of years, this formula is given:

$$\mu_{x+t} = -\frac{d}{dx} ln(l_{x+t}).$$

Again, due to the nature of human life spans, these values will be different for different survival functions, different countries, and there will be differences between the intervals researchers consider as areas of interest.

Next in the table is e^0_x. Recalling that this is the complete expectation of life at birth, a rough estimate for the life span of a life at age *x* can be calculated. Considering that this only takes into consideration the current age of the person, this is not the most precise portion of the graph, as well as it does not consider outside factors. This is generally used for "perfect health" expectations, so there are other methods used to more accurately determine expected life spans given certain health conditions or parameters, but regardless, this is another piece of information that can be useful. This is given by the following:

$$e^0_x = \int_0^\infty {}_tp_x dt.$$

This formula represents the expected lifespan of a person at age *x*, which can be used for either insurance purposes, or different studies that are conducted in the biological field. This formula can also be tailored to consider only a certain amount of years in the future. This formula is as follows:

$$e^0_{x:\overline{n}|} = \int_0^n {}_tp_x dt = \frac{1}{l_x}\int_0^n l_{x+t} dt.$$

This notation can be read as the expected value of the future lifetime of a life of age *x* over the next *n* years.

The following parts of the life table are L_x and T_x. These are not as easy to follow as the previous columns, but they can be understood fairly quickly. The website called “Life Tables for the United States Social Security Area 1900-2100” explains both of these parameters best. L_x is best thought as the number of people alive at their last age x birthday at any time in a stationary population (Felicitie, Miller 2). This is assumed to have a uniform distribution of deaths, which is where de Moivre’s law is applied in this sense. T_x is best thought of as the number of people alive at their last birthday x or older at any time in a stationary population. Just as the previous columns, these also have formulas that relate one another. These are:

$$L_x = l_x - \frac{1}{2}d_x \text{ and}$$

$$T_x = e^0_x \cdot l_x = L_x + L_{x+1} + L_{x+2} + \ .\ .\ .\ .$$

One thing to note about T_x is it is heavily related to e^0_x. The formula for e^0_x can also be defined as the moments generated by T_x. So,

$$E[T_x] = e^0_x$$

and the second moment and variance are:

$$E[T^2_x] = 2\int_0^\infty t \cdot \ {}_tp_x dt$$

$$Var[T_x] = 2\int_0^\infty (t \cdot \ {}_tp_x)dt - [\int_0^\infty \ {}_tp_x dt]^2.$$

These are mainly used to have more information about a population available, so despite not being the most relevant information, it is still useful. These could be used to compare differences of these values between populations and regions, showing disparities between any comparison that can be made. It does not even have to apply to live beings, as this could be applied in technological fields too. Comparing machine parts in as many ways as possible could sway the opinion of what piece of equipment should be used compared to another.

All of this information available to researchers can now be applied to any type of research field, insurance, biological fields, etc. The life table, now fully constructed and understood, plays a vital part in many different research projects that are conducted today. As the advancements in technology and medicine advance, these tables will continue to change. Current life tables will start to become obsolete, assuming advancements are made at the same pace as they are today, and newer life tables will always be needed. Hopefully these life tables allow for a better understanding of the world and all who inhabit it, with more researchers, funding, and experience, these newer models may come sooner than anticipated, with some studies already showing that may be the case.

Real-World Applications

Cox Proportional Hazard Model

As discussed earlier, there are many applications for these actuarial models. They have a wide range of use, and with that, they are seen in varying fields with different definitions based on these fields. The first example of these applications is the Cox Proportional Hazard Model and the approaches for testing the proportional hazard (PH) assumption.

The Cox Proportional Hazard Model is, in short, a regression model that is used for medicinal purposes. This model examines the survival times of patients with different variables depending on the treatments or drugs given. This article written about this specific hazard model, which was written by Frank Emmert-Streib and Matthias Dehmer, dives into how useful this model is when comparing cancer treatments and things of that nature. Before covering the finer details of this article, one section that is highly integral to medical trials and research, is censoring.

In the medical field, censoring is used when the researchers need to record the time of an event, but for varying reasons, only partial information will be obtained from a patient. This could happen with a patient that drops out of a study, withdraws from the study, or due to the limited time, the sought-after event does not occur until after the cut-off time. No matter the cause, these are all cases where censored data is used to lessen the effects of outliers or heavy tails that could pop up in research. An example of censoring is provided below.

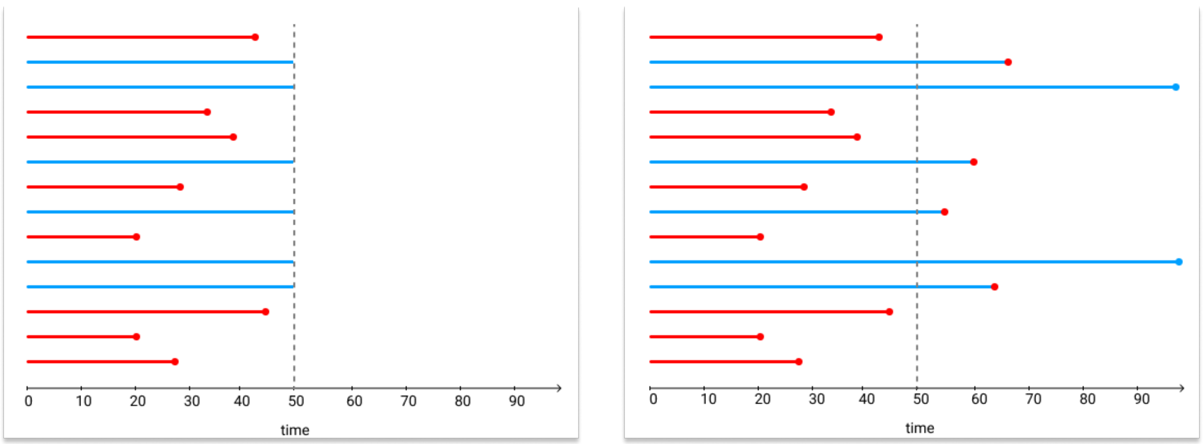

Figure 7. Censoring Graphs (*SB Professional Services*)

As shown, the graph on the left is what would be reviewed at the end of a study, with the events being cut off at the end time of the study. The graph on the right shows the true results of this study, which while it can be useful when reviewing studies and possible ways to change the studies, the researchers are less likely to include this information as it may misconstrued some data points one way or the other.

Next covered in the article is survival functions, hazard functions, and some of the models that are used. The main models used in this study are the Weibull, Exponential, Log-Normal, and a non-parametric model called the Kaplan-Meier (KM) Estimator for the Survival Function. The KM estimator is used to keep track of the number of subjects that experience the desired event, which makes sense as the numbers vary between each test, and this non-parametric model allows researchers to utilize the data discovered, since it would not always fit into one of the models discussed above.

The article then continues to discuss how the above models are used in their research. The hazard functions of each survival model all have different characteristics, which makes them useful in different scenarios. The article lists each of the models used, as well as the hazard function behavior and how it is used. For the Weibull model, it has the following hazard function:

$$h(t) = \lambda p(\lambda t)^{p-1} \text{ where } \lambda > 0 \text{ and } p > 0.$$

In this equation, λ is a rate parameters, and p is a shape parameter, so depending on the behavior of these parameters, the hazard functions can look very different. Thus, when $p = 1$, the hazard function is a constant, so this model is used for a normal, healthy patient. If $p = 2$, then the hazard function behavior becomes monotonous increasing. The usage then is for an unsuccessful surgery or treatment. The other model that is given a function is the Log-normal model. The hazard function for the Log-normal model is as follows:

$$h(t) = \frac{\alpha}{\sqrt{2\pi t}} e^{\frac{-\alpha^2 (ln(\lambda t))^2}{2}} (1 - \Phi(\alpha ln(\lambda t)))^{-1} \text{, where}$$

$$\alpha = \sigma \text{ of } ln(x) \text{, and}$$

$$\mu =- ln(\lambda).$$

With this model, the hazard function behavior is described as "humped" or "U-shaped," so this model is used for patients infected with tuberculosis or heart transplants.

All of these pieces are then placed together in order to form the Cox Proportional Hazard Model (CPHM). This model is used to include covariates of certain subjects, which the previous models could not take into consideration. The CPHM is a regression model with the following hazard function:

$$h(t, X) = h_0(t)e^{\beta_1 X},$$

where $h_o(t)$ is the baseline hazard. Some examples of when this is needed are given in the article, which include gender, smoking habits, or medication taken. One of the main uses for this model is to compare treatment groups with hazard groups. These groups are placed as $\frac{h(t, X^{treatment})}{h(t, X^{control})}$, and the value given from this equation can let the researchers know valuable information about which group experiences higher hazards compared to one another. These values can then be modeled over different tests to have a fully complete report that can be studied and published by many other researchers.

Overall, this is just one of many examples of how survival models can be used in an applicable setting. The Cox Proportional Hazard Model has many uses in the medical field, and it allows for researchers to compare their results with other studies, and more accurate information can be shared and studied.

Nelson-Aalen Estimator

Another commonly used application for survival models involves the use of the Nelson-Aalen estimator. This estimator is nonparametric, and an article by Ornulf Borgan states, "...may be used to estimate the cumulative hazard rate function from censored survival data." This is similar to the Kaplan-Meier (KM) estimator that was explained in the Cox Proportional Hazard model, but instead of tracking the number of subjects, the Nelson-Aalen Estimator is used to check the fit of certain parametric models with the use of graphs.

The first part of the Nelson-Aalen estimator that will be discussed is its application in survival data. This is the main component of this estimator, and all of the additions that have been made over the years require this first section to hold. Similar to the KM estimator, the data used for Nelson-Aalen needs to be censored data. With the function $A(t) = \int_0^t \alpha(s)\, ds$, this allows researchers to have an accurate record of data to be fitted. This data is then fitted using the following hazard rate function:

$$\hat{A}(t) = \sum_{t_j} \frac{d_j}{r_j},$$

where r_j is the number of individuals at risk just prior to time t_j (Borgan 1). Based on this model, this estimator can be seen as an increasing right-continuous step function with every step being $\frac{d_j}{r_j}$. Since this model is a summation, with large enough sample data, this model can be approximately normally distributed. This property can be used when checking the fits of models to see how close to normality the population studied is, given there is a large enough sample size. This step can be used to estimate the desired survival distribution function. The formula is:

$$S(t) = e^{-A(t)}.$$

This formula represents the probability that an individual will be alive at time *t*. This can be determined using the properties of the KM estimator. While the relation between the survival distribution using the properties of the KM estimator is incredibly useful for estimating the survival distribution, this is the only application for this specific model. There are, however, many more applications using just the Nelson-Aalen estimator that will be discussed. In the article, it is shown that there are two other major uses of the Nelson-Aalen estimator, these include relative mortality and epidemic models.

For relative mortality, the only assumption that is needed is the hazard rate function of the *i*th person can be written as $\alpha(t) \cdot \mu_i(t)$, where, as Borgan describes, "...$\alpha(t)$ is a relative mortality common to all individuals, and $\mu_i(t)$ is the hazard rate function at time *t* for a person from an external standard population corresponding to the *i*th individual (e.g. of the same sex and age as individual *i*)," (3). The $\mu_i(t)$ is typically known from already established life tables, which was defined and discussed above, so using this information, researchers are able to estimate cumulative relative mortality, which is:

$$A(t) = \int_0^t \alpha(s)\, ds.$$

The only requirement for this model is that r_j (recall this is the number of individuals at risk) should be used to denote the sum of the external rates for $\mu_i(t_j)$ for every individuals at risk prior to t_j. This is just one of the two ways that the Nelson-Aalen estimator is being used to estimate the cumulative relative mortality function.

In epidemic models, this is a relatively simple model that deals with infectious diseases. In this scenario, it is assumed that some individuals will come in contact that are infected with a certain disease, and they themselves become infected. This model deals purely with one person becoming infected from outside of a certain population, and no other infections will occur from outside of the population over the entire course of the epidemic. Thus, assuming that mixing in the community is random, the intensity of the infection is represented as $\alpha(t) \cdot S(t) \cdot I(t)$, where $S(t)$ is the number of susceptible people, $I(t)$ is the number of infected people, and $\alpha(t)$ is the infection rate. Due to this formula, the cumulative infection rate, $A(t)$, can be estimated by the Nelson-Aalen estimator where $r_j = S(t_j)I(t_j)$. Both of these examples show intricate use of the Nelson-Aalen estimator when determining $A(t)$, and these both incorporate many different survival models and can be curated to fit most survival models of interest to these researchers.

These uses of survival models are incredibly useful in today's world. Researchers can use the information that has been curated and studied for decades to have the most accurate information they can when performing research. The above examples are just a few of many applications of survival models used in the outside world. As time continues forward, there may be more applications for the models, and perhaps new uses will be discovered in just a short time. Despite the heavy use in insurance, these examples show the growing need in other fields to determine life-time data, not just for research purposes, but for the desire of knowledge that is so sought after in society today.

Conclusion

Actuarial science has a long developmental history with many different historical figures piecing together the entire field. From just the basics of developing probability, to the growing interest in mortality rates, this field has grown exponentially over the last hundred years. Many researchers have become more and more interested in survival models, and actuarial science as a whole. Countless years of research and history have helped develop this field from just the simple act of insuring ships and businesses involved in trade, to making real world impact in other fields

Each of the survival models created all have their own purpose and place in research today, with more applications being developed today. There is such a large well of untapped knowledge that is waiting to be explored with survival models. The applications that were explored above, while only a few, are representative of all the opportunities that could become of this information. As more people become interested in life-time rates, only time will tell all of the applications that researchers will find.

Works Cited

Arias-Castro, Ery. *Uniform Distribution Graph*. *Principles of Statistical Analysis: R Companion*, https://mathweb.ucsd.edu/~eariasca/R_companion_free/index.html. Accessed 15 Mar. 2025.

Bell, Felicitie C., and Michael L. Miller. "Life Table Methods." *The United States Social Security Administration*, www.ssa.gov/oact/NOTES/as116/as116_IV.html. Accessed 15 Mar. 2025.

Borgan, Ornulf. "Nelson–Aalen Estimator." *Wikipedia*, Wikimedia Foundation, 3 Feb. 2024, en.wikipedia.org/wiki/Nelson%E2%80%93Aalen_estimator.

Censoring Graph, https://i.loli.net/2020/11/21/qeHu4DydwamWB2Q.png. Accessed 15 Mar. 2025.

Coale, Ainsley J., and Paul Demeny. "Regional Model Life Tables and Stable Populations." *Google Books*, Google, books.google.com/books?id=7ESLBQAAQBAJ&lpg=PP1&ots=Hq5Vt2Khtg&dq=history+of+life+tables&lr&pg=PA3#v=onepage&q&f=false. Accessed 15 Mar. 2025.

Cunningham, Robin J., et al. Models for Quantifying Risk. ACTEX Publications, Inc, 2012.

Dong, Shiqi, *Analysis of the Variable Life Insurance Based on Log-Normal Distribution*, https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=02fadc75c0d2041fc734fdab8f367e8c3597afb7. Accessed 20 Mar. 2025.

Emmert-Streib, Frank, and Matthias Dehmer. “Introduction to Survival Analysis in Practice.” MDPI, Multidisciplinary Digital Publishing Institute, 8 Sept. 2019, www.mdpi.com/2504-4990/1/3/58.

Exponential Distribution. *Exponential Distribution: Uses, Parameters & Examples*, https://statisticsbyjim.com/probability/exponential-distribution/. Accessed 6 Apr. 2025.

Haberman, Steven. *Principles of Actuarial Science,* *https://www.actuariayfinanzas.net/PDFA&F/PrinciplesofActuarialScience_Sherris2001.pdf*. Accessed 15 Mar. 2025

Kennedy, Joseph Camp Griffith. “The Origin and Progress of Statistics.” *Journal of the American Geographical and Statistical Society*, vol. 2, no. 1, 1860, pp. 92–120. *JSTOR*, http://www.jstor.org/stable/45476627. Accessed 26 Jan. 2025.

Lesson 3: Life Tables, *https://www.measureevaluation.org/resources/training/online-courses-and-resources/non-certificate-courses-and-mini-tutorials/multiple-decrement-life-tables/lesson-3.html*. Accessed 15 Mar. 2025.

Life Tables, *https://www.datafix.com.au/BASHing/2020-03-11.html*. Accessed 15 Mar. 2025.

Log-normal distribution, *https://docs.analytica.com/index.php?title=Log-normal_distribution*. Accessed 15 Mar. 2025

Menoncin, Francesco. *Survival probability from 25 till t with a Gompertz-Makeham mortality law, https://www.researchgate.net/figure/Survival-probability-from-25-till-t-with-a-Gompertz-Makeham-mortality-law_fig1_241574021*. Accessed 15 Mar. 2025

Mumpar-Victoria, Marie. *Living to 100 and beyond: Survival at Advanced Ages*, www.soa.org/globalassets/assets/files/resources/essays-monographs/2002-living-to-100/mono-2002-m-li-02-1-faig-reply.pdf. Accessed 20 Mar. 2025.

Norström, Fredrik. "The Gompertz-Makeham Distribution." *Umea University*, 1997.

Pavlovic, Maja. "Log-Normal Distribution - a Simple Explanation." Towards Data Science, towardsdatascience.com/log-normal-distribution-a-simple-explanation-7605864fb67c/. Accessed 20 Mar. 2025.

Rotar′, Vladimir I. *Actuarial Models: The Mathematics of Insurance*. CRC Press, 2015.

Stigler, Stephen M. "The History of Statistics." *Google Books*, Google, books.google.com/books?id=nWn9DwAAQBAJ&lpg=PA1&dq=History+of+statistics&lr&pg=PP1#v=onepage&q&f=false. Accessed 20 Mar. 2025.

Uniform Distribution. *Essential Probability*, https://www.math.uh.edu/~charles/ess_prob_IV/essential_prob_4.htm. Accessed 6 Apr. 2025.

Winegar, Claire. "The History of Actuarial Science." *The Actuary Magazine*, 10 May 2024, www.theactuarymagazine.org/the-history-of-actuarial-science/.