6-2020

Research In Short Term Actuarial Modeling

Elijah Howells

Research In Short Term Actuarial Modeling

A Thesis

Presented to the

Faculty of

California State University,

San Bernardino

In Partial Fulfillment

of the Requirements for the Degree

Master of Arts

in

Mathematics

---

by

Elijah Howells

June 2020

Research In Short Term Actuarial Modeling

A Thesis

Presented to the

Faculty of

California State University,

San Bernardino

---

by

Elijah Howells

June 2020

Approved by:

Joseph Chavez, Committee Chair

Yuichiro Kakihara, Committee Member

Rolland Trapp, Committee Member

David Maynard, Chair, Department of Mathematics

Corey Dunn, Graduate Coordinator

Abstract

This paper covers mathematical methods used to conduct actuarial analysis in the short term, such as policy deductible analysis, maximum covered loss analysis, and mixtures of distributions. Assessment of a loss variable's distribution under the effect of a policy deductible, as well as one with an implemented maximum covered loss, and under both a policy deductible and maximum covered loss will also be covered. The derivation, meaning, and use of cost per loss and cost per payment will be discussed, as will those of an aggregate sum distribution, stop loss policy, and maximum likelihood estimation. For each topic, special cases based on distribution will be described and discussed. These methods and subjects are used to assess and manage risk, typically for insurance providers, but can also be adapted to a number of other fields.

Acknowledgements

I would like to thank Dr. Joseph Chavez for his guidance and aid in researching the material covered in this project, as well as Dr. Roland Trapp and Dr. Yuichiro Kakihara for their comments and critiques on this paper. I am very thankful for the assistance.

Table of Contents

Chapter 1

Introduction

1.1 Short Term Actuarial Models

An actuarial model is the application of a mathematical statistical distribution to a real-world, recurring situation. These will typically be in the realm of financial analysis, and aim to predict financial risks. These models are not expected to be perfectly accurate, and oftentimes prioritize functional understanding of risk involved in making financial decisions for the sake of minimizing unnecessary losses. Actuarial models that try to describe the happenings in a relatively short period of time, ranging anywhere from a few days to any period less than 5 years, are called Short Term Models. These will deal more predominantly with individual payouts or the number of payments to be made per month or year, and are meant to deal with smaller probability spaces in order to more accurately understand what is happening with some specific policy or event.

1.2 Methodology

In mathematical statistics it is commonplace to experiment with and analyze situations wherein a random variable is modified by adding, subtracting, or multiplying the variable within its distribution functions by some fixed amount. This can give an understanding of how the variable's distribution can be impacted by changes in its probability space, such as restricting the values to the rationals, integers, etc. The foundations of Short Term Actuarial Modeling can be best described as drawing connections between these types of modifications to the random variable, or its probability space, and real-world changes made to a recurring event. This is then expanded upon by looking at possible changes to the real-world event and quantifying them in ways that combine alterations to the variable for the sake of understanding how a less obvious change in the distribution can be brought about. The goal of this is to more accurately predict how the implementation of changes in an insurance policy deductible, maximum covered loss, or portion of damage covered will impact average payouts. There may be times when a policy can't be accurately described by a single distribution, and we must consider a mixture of two or more distributions in order to properly assess the policy's future costs.

We then extrapolate from that the total amount paid out to all members covered by that policy, and can then use that to project net profits. With these fundamentals understood, we can then begin applying them to less obvious scenarios. What are our options for reducing payouts by some amount? If we know that this policy's payments follow a specific distribution, but the individual payments observed have changed in amount over time, have our distribution's parameters changed?

1.3 What is Covered in This Project

This project will begin with a summary of the fundamental methods to be applied to more advanced modeling techniques. The focus of this section will be on deductibles and maximum covered losses, which will come up when discussing nearly every other topic. The use of mixed distributions and operations performed with them will also be discussed. Special cases of deductibles, maximum covered losses, and mixtures will also be covered for the purpose of making later work easier.

In the next section more advanced methods of actuarial analysis, and how they are affected by deductibles and maximum covered losses will be addressed, will be explored. Aggregate payments as a combination of a severity distribution and a frequency distribution will be studied, as well as how an aggregate model is impacted by the introduction of a deductible and/or maximum covered loss on the severity. The method of convolution and its use in the derivation of stop-loss premiums will also be covered. Maximum likelihood estimation, the derivation of likely parameters of a given distribution based on an observed data set, will be explored as well. Finally, special cases of aggregate models and maximum likelihood estimation will be outlined for the sake of making these processes easier in the future.

1.4 Applications and Adaptations

To conclude, we will apply the methods covered in this project to real-world scenarios based on data collected from publicly available sources; this direct application will pertain mainly to insurance claims predictions. Much of this application will be straightforward, since insurance is the context in which the modeling methods were developed. However, we will also show how actuarial modeling can be used to predict damages from a natural disaster, such as an earthquake striking in an urban population center.

Chapter 2

Summary of Methods

The Society of Actuaries (SOA) is an organization that acts as the main accreditation board for actuaries in the United States. SOA accreditation is broken into three main classes, each with different exams, courses, and research required. However, one common requirement for all classes of accreditation is passing the actuarial exam on Short Term Actuarial Modeling (STAM), sometimes called Short Term Actuarial Mathematics.

To aid with preparation for the exams, the SOA published books and manuals covering the material that one will need to know in order to pass the exams. Two such texts are Loss Models and the ACTEX Exam STAM Study Manual, which both deal with material used on the STAM exam. The Loss Models text is considered to be one of the definitive sources for contextualizing and explaining short term actuarial modeling. On the other hand, the ACTEX manual is something of an exam study guide that compiles and distills all the relevant equations and formulae, and contains a number of exercises and practice exams to study from.

In this chapter, we will be covering the fundamental methods that are important for real-world use and passing the exam. These methods are the basis for some of the more advanced ones that will come up later, and are among the most commonly utilized in real-world actuarial work. Due to the nature of this exposition, the two aforementioned texts are vital since they bring together all the necessary information in a well organized fashion.

2.1 Maximum Covered Loss

When we use an unmodified random variable to represent losses, we call it the **ground up loss variable**. It is typical to modify this variable to reflect some kind of loss limiting method being implemented. One of the most common policy modifications in the insurance industry is the application of a **maximum covered loss (mcl)**, a set upper limit on the amount for a single insurance claim. Assuming no other modifications are in place, an mcl can be fairly simple to work with since it just acts as a maximum value for the loss variable. Typically denoted u, the mcl can be combined with the ground up loss variable, X, to form the limited loss random variable. Formally, we define this in the following way.

Definition 2.1. *[Bro19]*

$$\textbf{\textit{Limited Loss Random Variable}} = X \wedge u = \begin{cases} x \ \textit{if} \ x < u \\ u \ \textit{if} \ x > u \end{cases}$$

In order to help make the writing cleaner when working with an mcl, we can set $Y = X \wedge u$, and use this as our new random variable. This allows us to describe the distribution for the limited loss variable in terms of the ground up loss variable, including a density function for Y.

Definition 2.2. *[Bro19] The* ***pdf of Limited Loss Variable Y*** *is a shift of the ground up loss variable's pdf by a factor of u.*

$$f_Y(y) = \begin{cases} f_X(y) & \textit{if} \ y < u \\ 1 - F_X(y) & \textit{if} \ y = u \\ 0 & \textit{if} \ y > u \end{cases}$$

The second line comes from the fact that $F_X(y)$ when $y = u$ gives the probability that $X < u$ before applying the maximum value u to X. So, $1 - F_X(y)$ is the best approximation of the likelihood of u occurring prior to assuming that u is the maximum value. From the pdf of Y we can also derive the distribution function of Y.

$$F_Y(y) = \begin{cases} F_X(y) & if\ y \leq u \\ 1 & if\ y > u \end{cases}$$

Both functions work and behave the same way that they would for a ground up loss variable, they just take into account the shift caused by limiting X. As such we can use them to find probabilities for the values of $Y = X \wedge u$ and the likelihood that Y will be less than or greater than a given value or between two given values. It is also worth noting that these shifted functions do not depend on whether X has a continuous or discrete distribution, and that Y will inherit distribution of X. Furthermore, we can use this in order to find the expected value of Y.

Theorem 2.3. *[Klu12] The expected value of Y is,*

$$E[Y] = E[X \wedge u] = \begin{cases} \left(\int_0^u xf(x)dx\right) + u[1 - F_X(u)] & if\ X\ is\ continuous \\ \left(\sum_{x_j \leq u} x_j p(x_j)\right) + u[1 - F_X(u)] & if\ X\ is\ discrete \\ \int_0^u xf_X(x)dx + u[1 - F_X(u)] & if\ X\ is\ either\ disc.\ or\ cont. \end{cases}$$

We can also split this integral into two ranges, $-\infty \leq x < 0$ and $0 \leq x \leq u$. However, we are typically not concerned with any $x < 0$ because our objective is to find the expected number of occurrences of an event or to anticipate the value of an expenditure. Since we cannot have a negative number of occurrences and we cannot have expenditures less than zero, we do not need to include the integral that covers the negative values of x. The end result is the following equation, which is notably easier to work with and just as accurate.

$$E[Y] = E[X \wedge u] = \int_0^u 1 - F_X(x)dx = \int_0^u S(x)dx \tag{2.1}$$

Example 2.1.1. *Consider the ground-up loss random variable X with exponential distribution, representing the loss amount on an insurance claim for car repairs. If X has a mean of \$750 and a maximum covered loss of \$1500, then we can find the expected value of $(X \wedge 1500)$ by doing the following.*

$$E[X \wedge 1500] = \int_0^{1500} 1 - [1 - e^{-x/750}]dx = \int_0^{1500} e^{-x/750}dx = 750 - 750e^{-2} = \boldsymbol{648.50}$$

This example also demonstrates why the simplified version of the expected value equation is so much more convenient to use than the original. However the original version is still useful, as it reflects the way of calculating any moment of Y. To find the $kth - moment\ of\ X \wedge u$ for a continuous X.

Theorem 2.4. *[Klu12] The kth-moment of Y is,*

$$E[Y^k] = E[X \wedge u] = \int_0^u x^k f_X(x)dx + u^k[1 - F(u)]$$

With this, we can find the variance of Y and by extension it's standard deviation. There is no shortcut to finding this, so we go back to the variance's definition.

$$Var[X \wedge u] = E[(X \wedge u)^2] - (E[X \wedge u])^2$$

2.2 Deductibles

Another common modification made to insurance policies is the introduction of a deductible. There are two main types of deductibles that may be implemented, and the mathematics associated with them is similar but distinct. The first type is called an **Ordinary Deductible**, or a policy deductible, and acts exactly as the name implies: a fixed deduction from all claims made under the given policy. The other type however, a **Franchise Deductible**, works more like a minimum claim size. Both are common and useful tools when making predictions for a given policy or assessing risk. Whether we are using one type or the other, the main discussion will be focused on the cost per loss of the given ground up loss variable.

Definition 2.5. *[Bro19] The* ***cost per loss*** *denoted by Y_L, of a given random variable is the total amount that an insurer pays on losses represented by that variable after all policy modifications have been applied to those losses.*

In this section, we will be looking specifically at deductibles as policy modifications. The typical notation is to let d be the deductible, which gives us a cost per loss of $Y_L = (X - d)$.

2.2.1 Ordinary Deductibles

Most of the time a deductible appears, it is an ordinary deductible on some kind of insurance policy. In these cases, the focus is place predominantly on understanding the cost per loss, Y_L, and what to expect of it. We can define Y_L more specifically as follows.

$$Y_L = (X-d)_+ = \begin{cases} 0 & if\ X \leq d \\ X-d & if\ X > d \end{cases} = max(X-d, 0) \tag{2.2}$$

Notice that $X = (X \wedge d) + (X-d)_+$, because if $X \leq d$ then $(X \wedge d) = X$ and $(X-d)_+ = 0$; and if $X > d$ then $(X \wedge d) = u$ and $(X-d)_+ = (X-d)$. In both cases, the sum of the two sets is just X. From this we have that $Y_L = (X-d)_+ = X - (X \wedge d)$, which is a very useful way of expressing Y_L. Using this we can write the pdf and cdf of the cost per loss variable Y_L.

Definition 2.6. *[Klu12] The* ***pdf of*** Y_L *is given by the piecewise function,*

$$f_{Y_L}(y) = \begin{cases} F_X(d) & if\ y = 0 \\ f_X(y+d) & if\ y > 0 \end{cases}$$

From this definition of the pdf, the ***cdf of*** Y_L *is reasonably apparent.*

$$F_{Y_L}(y) = \begin{cases} F_X(d) & if\ y = 0 \\ F_X(y+d) & if\ y > 0 \end{cases}$$

With these functions we can calculate the expected value of Y_L using the definition and finding the first moment, but that can become cumbersome very quickly. Fortunately, we have the definition of $Y_L = X - (X \wedge d)$ to make things easier. Since all functions being used here are linear, and finding the first moment of a distribution is also linear, we can say that $E[Y_L] = E[X] - E[X \wedge d]$. After applying previously established methods for finding these values and simplifying, we get the following equation.

Theorem 2.7. *[Bro19] The expected value of* Y_L *when* d *is an ordinary deductible is,*

$$E[Y_L] = E[(X-d)_+] = E[X] - E[X \wedge d] = \int_d^\infty [1 - F_X(x)]dx = \int_d^\infty S_X(x)dx$$

Example 2.2.1. *Consider the ground-up loss random variable X with exponential distribution and a mean of \$750 and an ordinary deductible of \$250. The expected value of X will be \$537.40, as shown below.*

$$E[X-250] = \int_{250}^{\infty} 1-(1-e^{-x/750})dx = \int_{250}^{\infty} e^{-x/750}dx = 750(e^{-250/750}) = \boldsymbol{537.40}$$

The variance of Y_L is not quite as conveniently derived though, since it relies on the second moment of Y_L which is not linear. That is, we cannot say that $E[Y_L^2] = E[X] - E[X \wedge d]$. Rather, after a bit of work we arrive at the second moment of Y_L being the following,

$$E[Y_L^2] = E[X^2] - E[(X \wedge d)^2] - 2d(E[X] - E[X \wedge d])$$

Now we use the definition of the variance and the shortcut to finding $E[Y_L]$ to find the variance of Y_L.

$$\begin{aligned} Var[Y_L] &= E[Y_L^2] - (E[Y_L])^2 \\ &= E[X^2] - E[(X \wedge d)^2] - 2d(E[X] - E[X \wedge d]) - \Big(E[X] - E[X \wedge d]\Big)^2 \\ &= E[X^2] - E[(X \wedge d)^2] - 2d(E[X] - E[X \wedge d]) - \left(\int_d^{\infty} [1 - F_X(x)]dx\right)^2 \end{aligned} \tag{2.3}$$

Another value that we may want to know is the **Loss Elimination Ratio** (LER) of X with a deductible d. The LER is the expected amount of money saved by an insurer if they implement a deductible of d on a policy with claim sizes from X. Naturally, this value is something that can be very useful when trying to compare two or more potential policies. The LER can be derived by dividing the expected value of $(X \wedge d)$ by the expected value of X. In notation:

$$LER(X-d) = \frac{E[X \wedge d]}{E[X]} \tag{2.4}$$

Our final concern with and ordinary deductible, is the value of losses given a loss is greater than the deductible. This is called the **Cost Per Payment of** X, and is expressed as a conditional value of $X - d$.

$$Y_P = \frac{Y_L}{P(X>d)} = \frac{(X-d)}{P(X>d)} \tag{2.5}$$

Typically, we are only really concerned with the the expected value and variance of Y_P. They can be found fairly easily thanks to the linearity of the expected value definition and the close relationship with Y_L.

$$E[Y_P] = \frac{E[Y_L]}{1 - F_X(d)} = \frac{\int_d^\infty 1 - F_X(x)dx}{1 - F_X(d)} \tag{2.6}$$

$$Var[Y_P] = \frac{E[Y_L^2]}{1 - F_X(d)} - \left(\frac{E[Y_L]}{1 - F_X(d)}\right)^2 \tag{2.7}$$

Unfortunately since there is no particularly short way of deriving the second moment of Y_L, we can't come up with a shorter way of finding variance of Y_P.

2.2.2 Franchise Deductibles

The second type of deductible that can be applied is the franchise deductible. It behaves similar to an ordinary deductible, with the main difference being that when $X > d$ the insurer covers the entire claim, rather than the claim minus the deductible's amount. This acts as a kind of pdf for the amount paid, and can be thought of as a modification on the lower limit of X. We can consider the loss amount to be X with a lower limit of d rather than 0. In this context, we define Y_L as follows.

$$Y_L = \begin{cases} 0 & if\ X \leq d \\ X & if\ X > d \end{cases} \tag{2.8}$$

The method of finding the expected amount paid by the insurer when a franchise deductible is applied is essentially using the definition of the expected value of X, but taking into account the fact that d behaves as a new lower bound for X.

Theorem 2.8. *[Klu12] The expected value of Y_L when X is subject to an franchise deductible of d is,*

$$E[Y_L] = \int_d^\infty x f_X(x)dx$$

We can also put this in terms of d being an ordinary deductible. To do this we only need to notice that in this case,

$$Y_L = \begin{cases} 0 & if\ X \leq d \\ (X-d)+d & if\ X > d \end{cases}$$

Using this form of Y_L, we can use the expected value of $(X-d)$ as a basis for the expected value of Y_L. However, we have to add to it the expected portion of d that will appear. To satisfy this we can add $d[1-F_X(d)]$, since $[1-F_X(d)]$ is the probability that $X > d$, the product of this value and d itself will be the portion of d to expect in non-zero losses. With this, we get the following alternative form of $E[Y_L]$.

$$E[Y_L] = E[X-d] + d[1-F_X(x)]$$

Example 2.2.2. *Now let's revisit Example 2.2.1, but instead of an ordinary deductible X will have a franchise deductible. So, consider X, an exponential ground-up loss variable, with a mean of \$750 and a franchise deductible of \$250. We find the expected value of X after the deductible is applied to be \$716.53, by doing the following.*

$$\int_{250}^{\infty} \frac{x}{750} e^{-x/750} dx = 537.40 + 179.13 = \boldsymbol{716.53}$$

2.3 Combination of Max Covered Loss and Deductible, and Cost Per Payment

In the insurance industry it is normal for a policy to include both a deductible and a maximum covered loss. From here on, we will assume that all deductibles discussed are ordinary, unless specified otherwise. When dealing with both, it is important to understand what the policy's limit is.

Definition 2.9. *[Klu12] A **policy limit** is the maximum amount that an insurer will pay on a given policy, per claim, after all policy modifications are applied.*

In this case, we must consider which of the two modifications is applied first. If the deductible is applied before the maximum covered loss then the policy limit is equal to the max covered loss; however if the maximum covered loss is applied first, then the policy limit is the max covered loss minus the deductible. Combining both a max covered loss, u, and deductible, d, with the ground-up loss variable X, we can find the **cost per loss,** Y_L. Let's assume that u is applied before d to outline Y_L.

$$Y_L = \begin{cases} 0 & if\ X \leq d \\ X - d & if\ d < X \leq u \\ u - d & if\ X > u \end{cases} \tag{2.9}$$

A useful observation to make here is that we are effectively restricting the interval that X is defined one. The deductible acts as a lower bound and the mcl acts as an upper bound. Another way to look at this though, is to think of it as the interval $[0, u] - [0, d]$. Viewing it this way let's us treat Y_L as a difference between two max covered losses being placed on X.

$$Y_L = (X \wedge u) - (X \wedge d) \tag{2.10}$$

Furthermore, we can find the pdf and cdf of Y_L.

Theorem 2.10. *[Bro19] The distribution functions of Y_L are derived from those of X.*

$$f_{Y_L}(y) = \begin{cases} F_X(d) & if\ y = 0 \\ f_X(y + d) & if\ 0 < y < u - d \\ 1 - F_X(u) & if\ y \geq u - d \end{cases}$$

$$F_{Y_L}(y) = \begin{cases} F_X(d) & if\ y = 0 \\ F_X(y + d) & if\ 0 < y < u - d \\ 1 & if\ y \geq u - d \end{cases}$$

This is the case that we typically concern ourselves with, but the case where the deductible is applied first also occurs occasionally. This case works nearly the same, but in the last line we would change the bound on y from $u - d$ to u. Giving us the last line, $Y_L = u$ if $X > u$. The difference between the max covered loss and the policy limit becomes more important when discussing policies with others and trying to compare policies. If we are talking about a policy with an mcl applied before the deductible and say, “This policy has a deductible of \$300 and a policy limit of \$500,” then we’re really saying that $u = \$800$, which has a much different impact on calculating the expected value of Y_L than using $u = \$500$. The expected value of the cost per loss, $E[Y_L]$, also called the **expected cost per loss** is fairly straightforward. Since the expected value is a linear function, we can just distribute the E across $(X \wedge u) - (X \wedge d)$. This can also be rewritten as a single integral evaluated from d to u.

Theorem 2.11. *[Klu12] The expected cost per loss when both an ordinary deductible and maximum covered loss are present is,*

$$\begin{aligned} E[Y_L] &= E[X \wedge u] - E[X \wedge d] \\ &= \int_d^u [1 - F_X(x)]dx = \int_d^u S(x)dx \end{aligned}$$

There is one alteration we should consider with this formula though. Here we are assuming d to be an ordinary deductible, but what if it is a franchise deductible instead? When this is the case, we need to use a modified version of the first line and the formula would change.

$$\begin{aligned} E[Y_L] &= E[X \wedge u] - E[X \wedge d] + d[1 - F_X(d)] \\ &= \int_d^u x f_X(x)dx + d[1 - F_X(d)] \end{aligned}$$

These functions mirror those for expected values of random variables with deductibles, just changing the upper bound from infinity to u. Now let’s return to the example described above to see the typical way we would use this.

Example 2.3.1. *Assume X to be the ground-up loss variable with exponential distribution, a mean of \$750, a deductible of \$300, and a policy limit of \$500. Assuming that the mcl is applied before the deductible, this means that $u = \$800$, and our expected value is as follows.*

$$E[Y_L] = \int_{300}^{800} e^{-x/750}dx = \boldsymbol{244.62}$$

However if the mcl is applied *after* the deductible then the expected value is \$117.68, considerably different from the previous result. This discrepancy becomes much more impactful when considering that an insurance provider will have thousands of customers with the same type of policy, and as a result we would see the \$126.94 difference being multiplied by however many customers purchase that policy. As stated before, normally u is applied first and our policy limit is $u-d$ so unless specified otherwise the following explanations will reflect this ordering, but in practice it is important to double check the ordering of modifications. As is typical with calculations like these, the difference between these two interpretations becomes even more drastic when looking at the second moment of Y_L. It's easy enough to anticipate this trend as a mathematician, since the second moment is a function of degree two as opposed the to first moment being of degree one. Finding the second moment can be cumbersome though, as there is no particularly concise way of doing it.

$$\begin{aligned} E[Y_L^2] &= \int_d^u (x-d)^2 f_X(x)dx + (u-d)^2[1-F_X(u)] \\ &= \Big(E[(X\wedge u)^2] - E[(X\wedge d)^2]\Big) - 2d\Big(E[X\wedge u] - E[\wedge d]\Big) \end{aligned} \tag{2.11}$$

Another important concept in policy modifications is the Cost per Payment. It is similar to a cost per loss, but dismisses the number of losses that are less than the deductible. Intuitively, we aren't particularly concerned with the losses less than the deductible since the don't result in any loss "on our end," when in the mindset of an actuary.

Definition 2.12. *[Bro19] The **Cost per Payment**, denoted Y_P, of X with max covered loss u and deductible d, is the cost per loss divided by $P(X > d)$.*

$$Y_P = \frac{Y_L}{P(X>d)} = \frac{Y_L}{1-F_X(d)} \tag{2.12}$$

Now, we can find the pdf and cdf of Y_P similar to how we did with Y_L.

$$f_{Y_P}(y) = \begin{cases} \frac{f_X(y+d)}{1-F_X(d)} & if\ 0 < y < u-d \\ \frac{1-F_X(u)}{1-F_X} & if\ y \geq u-d \end{cases}$$

$$F_{Y_P}(y) = \begin{cases} \frac{F_X(y+d)-F_X(d)}{1-F_X(d)} & if\ 0 < y < u-d \\ 1 & if\ y \geq u-d \end{cases} \tag{2.13}$$

The expected value and variance of Y_P are fairly simple to find, but we must use the definition of the variance as there is no simplified form that works in the general Y_P case. Fortunately though, the moments of Y_P are just the corresponding moments of Y_L divided by $[1 - F_X(d)]$. We can just write this as follows.

$$E[Y_P^k] = \frac{E[Y_L^k]}{1 - F_X(d)}$$

Our main concern is with $E[Y_P]$ and $E[Y_P^2]$, since we need the second moment to find the variance of Y_P. Using the above formula, we get the following for the expected value and variance of Y_P.

$$E[Y_P] = \frac{E[Y_L]}{1 - F_X(d)} = \frac{E[X \wedge u] - E[X \wedge d]}{1 - F_X(d)} \tag{2.14}$$

$$Var[Y_P] = E[Y_P^2] - \big(E[Y_P]\big)^2 \tag{2.15}$$

2.4 Mixtures

Sometimes, a random variable may not be adequately represented by a single distribution. Instead we can consider the variable, Y, to be a **mixture** of two or more variables. These **component variables**, $X_1, X_2, ..., X_n$ each have their own distribution, and each contributes a different portion of Y. Each component variable, X_i, will have an associated **mixing weight**, typically denoted by α_i, which must satisfy two properties: $0 < \alpha_i < 1$ and $\sum_{i=1}^{n} \alpha_i = 1$. Using the component variables' distributions and mixing weights, we can derive a density function for the mixed variable Y.

Definition 2.13. *[Klu12] The* ***pdf of Mixed Variable*** *Y is the weighted average of the pdfs of its component variables.*

$$f_Y(y) = \alpha_1 f_{X_1}(y) + \alpha_2 f_{X_2}(y) + ... + \alpha_n f_{X_n}(y)$$

Using this, and the fact that all α_i's are constants it follows that the cdf of Y, which is just the integral of Y's pdf, is $F_Y(y) = \alpha_1 F_{X_1}(y) + \alpha_2 F_{X_2}(y) + ... + \alpha_n F_{X_n}(y)$. From here, we can derive various values and formulas for values related to Y's distribution. One of the most important, if not the most important, value we should look for is the kth-moment $of\ Y$.

Theorem 2.14. *[Bro19]*

$$E[Y^k] = \sum_{i=1}^{n} \alpha_i E[X_i^k]$$

In English, we can say that the kth-moment of Y is given by the weighted average of its component variables' kth-moments. This equation is used very frequently, as there is no concise way of deriving the variance or standard deviation of Y from those of its component variables. So in order to find Var$[Y]$ we must use the moment-based definition, $E[Y^2] - (E[Y])^2 = (\sum_{i=1}^{n} \alpha_i E[X_i^2]) - (\sum_{i=1}^{n} \alpha_i E[X_i])^2$. In spite of this, many of the important values and functions that relate to Y's distribution will actually be the weighted averages of their component counterparts. The main values that we care about with this property are probabilities of intervals, moment generating functions, and probability generating functions of Y, which all have fairly intuitive ways of being written as weighted averages of the component functions' respective values and functions. We can say that if Y is a mixture of n components, then:

$$P(a < Y < b) = \sum_{i=1}^{n} \alpha_i P(a < X_i < b) \tag{2.16}$$

$$M_Y(t) = \sum_{i=1}^{n} \alpha_i M_{X_i}(t) \tag{2.17}$$

$$P_Y(t) = \sum_{i=1}^{n} \alpha_i P_{X_i}(t) \tag{2.18}$$

Typically these methods and formulas are sufficient for working with a mixed distribution. However, they are not always the easiest ways of doing things. Particularly, we can use some general rules of probabilities to make dealing with mixtures of large numbers of distributions a bit easier. In that regard, there are two main principles we should consider.

Theorem 2.15. *[Klu12] We can rewrite the expected value and variance formulas as conditionals based on surrounding events or parameters of the variable's distribution.*

$$E[X] = E[E[X|W]]$$
$$Var[X] = Var[E[X|W]] + E[Var[X|W]]$$

These properties are useful whenever we deal with probabilities and statistics, but are especially useful when considering a mixture of distributions. This is because we can consider the distribution of Y to be conditional on the number of component variables being used. We do this by defining the **conditioning random variable** $\Theta = \{1, 2, ..., n\}$. Then we can write Y as a mixture of X_i's with the mixing weights $\alpha_i = P[\Theta = i]$. This gives the relation $f_{Y|\Theta}(y|\Theta = i) = f_{X_i}(y)$, however we can use the relation to write an unconditional pdf for Y.

Definition 2.16. *[Klu12] The **Unconditional pdf of Y** with conditioning random variable* $\Theta = \{1, 2, ..., n\}$ *is*

$$f_Y(y) = \sum_{i=1}^{n} \alpha_i f_{X_i}(y) = \sum_{i=1}^{n} f_{Y|\Theta}(y|\Theta = i)P(\Theta = i)$$

This definition relies on the distribution of Θ being discrete, but a continuous distribution on Θ is also possible. If we define Θ to be continuous, but have the same relationship with Y, then we get a similar unconditional pdf of Y that must be integrated rather than summed. That is,

$$f_Y(y) = \int f_{Y|\Theta}(y|\Theta) f_\Theta(\theta) d\theta \tag{2.19}$$

The limits of this integral are the bounds of θ. Since our continuous variables are usually in the interval $(0, \infty)$, that is a common interval for $f_Y(y)$. Now we can use this to find formulae for the typical values we concern ourselves with.

$$E[Y] = \int E[Y|\Theta = \theta] f_\Theta(\theta) d\theta \tag{2.20}$$

$$E[Y^k] = \int E[Y^k|\Theta = \theta] f_\Theta(\theta) \tag{2.21}$$

[Bro19]

$$F_Y(y) = \int F_{Y|\Theta}(y|\theta) f_\Theta(\theta) d\theta \tag{2.22}$$

$$P(a \leq Y \leq b) = \int P(a \leq Y \leq b|\Theta = \theta) f_\Theta(\theta) d\theta \tag{2.23}$$

This is where it becomes exceptionally helpful to use the two principles defined previously, $E[X] = E[E[X|W]]$ and $Var[X] = Var[E[X|W]] + E[Var[X|W]]$, as these can reduce the amount of work involved in finding $E[Y]$ and $Var[Y]$ when most information is given conditionally.

2.5 Special Cases

Through luck and happenstance, certain distributions work very well with different modifications and mixtures in ways that result in drastically reduced amounts of work after cancellation in general forms. Some of these special instances are noteworthy, either because they can make certain processes into a small fraction of the work they would be otherwise, or because they pertain to distributions that are exceedingly common and will be used often in application. The two main areas where special cases can be useful are special cases of deductibles and special cases of mixtures.

2.5.1 Deductibles of the Exponential and Pareto Distributions

Our primary concern with these cases will be the cost per payment, as this is where the work involved becomes most cumbersome. However because of the added levels of complexity involved, cancellation can occur more often than in other coverage models. There are some ways to reduce nearly all distributions' cost per loss, but the exponential distribution has a particularly interesting property that helps to reduce its cost per loss more than any other distribution, and the Pareto distributions, both type 1 and type

2, are some of the most commonly used distributions in application and are reasonably convenient in their formula reductions.

The exponential distribution has a useful property called "lack of memory," which means that when a condition is placed on the exponential variable, the conditional probabilities will be equal to the non-conditional probabilities. When a modification is placed on the loss variable, the new distribution is still going to be exponential and typically with the same mean, just with a different representative variable being used. A good example of this, and the one that we are predominantly concerned with here, is the distribution of Y_P when X is exponentially distributed and has an ordinary deductible of d applied.

Theorem 2.17. *[Klu12]*

$$\begin{aligned} f_{Y_P}(y) &= \frac{f_X(y+d)}{1-F_X(d)} = \frac{\frac{1}{\theta}e^{-(y+d)/\theta}}{e^{-d/\theta}} \\ &= \frac{1}{\theta}e^{-y/\theta} \end{aligned}$$

This is another exponential distribution with the same mean, but instead of the loss variable X, we have $Y = (X - d)$. Because of this convenience. we can also use the typical formulae associated with the moments and variance of an exponential variable. Which means that $E[Y_P] = \theta$, $Var[Y_P] = \theta^2$, and $E[Y_P^k] = (k!)\theta^k$.

Similarly, if X is a ground-up loss variable with Pareto type 2 distribution and the parameters α and θ, then we can find the pdf of Y_P as follows.

$$f_{Y_P}(y) = \frac{f_X(y+d)}{1-F_X(d)} = \frac{\alpha\theta^\alpha}{(y+d+\theta)^{\alpha+1}} \Bigg/ \left(\frac{\theta}{d+\theta}\right)^\alpha$$

$$= \frac{\alpha(\theta+d)^\alpha}{(y+d+\theta)^{\alpha+1}} \; for \; y \geq 0$$

Replacing $\theta + d$ with θ_{Y_P}, we get the new Pareto type 2 distribution:

$$f_{Y_P}(y) = \frac{\alpha\theta_{Y_P}^\alpha}{(y+\theta_{Y_P})^{\alpha+1}} \tag{2.24}$$

which has the same α as X, and maintains all the formulae associated with a standard type 2 Pareto distribution. Thus, we have that $E[Y_P] = \frac{\theta_{Y_P}}{(\alpha-1)} = \frac{(\theta+d)}{(\alpha-1)}$,
$Var[Y_P] = \frac{2(\theta+d)^2}{(\alpha-1)(\alpha-2)} - \left(\frac{\theta+d}{(\alpha-1)}\right)^2 = \frac{\alpha(\theta+d)^2}{(\alpha-1)^2(\alpha-2)}$, and $E[Y_P^k] = \frac{(\theta+d)^k k!}{(\alpha-1)...(\alpha-k)}$.

Pareto type 1 is particularly special when a deductible is applied. A variable, X, is only defined for all $X > \theta$, meaning that if $d \leq \theta$ then X is unchanged and all values remain the exact same. So really, we are only concerned with the $d > \theta$ case. Here, we can expand the pdf of Y_P to arrive at an interesting, and very convenient, conclusion.

$$f_{Y_P}(y) = \frac{f_X(y+d)}{1-F_X(d)} = \frac{\alpha\theta^\alpha}{(y+d)^{\alpha+1}} \Bigg/ \left(\frac{\theta}{d}\right)^\alpha = \frac{\alpha d^\alpha}{(y+d)^{\alpha+1}} \; for \; y \geq 0 \tag{2.25}$$

This final pdf is actually that of a Pareto type 2 variable with the same α as X and $\theta = d$. As such, the expected value, moments, and variance of Y_P follow the standard formulas for those of a Pareto type 2 distribution. So, $E[Y_P] = \frac{d}{\alpha-1}$, $Var[Y_P] = \frac{\alpha d^2}{(\alpha-1)^2(\alpha-2)}$, and $E[Y_P^k] = \frac{d^k k!}{(\alpha-1)...(\alpha-k)}$.

2.5.2 Special Distribution Mixtures

Depending on the distributions of the component variables of a mixture, there some things that we can say about the distribution of the mixed variable itself. For instance if all component variables, X_i, are discrete, then the mixture, Y, will be discrete as well. Likewise, if all X_i are continuous then Y will be continuous. However if some components are discrete and some are continuous, we cannot say much about the mixture as it could be discrete, continuous, or a mixed distribution.

Similar to a mixed distribution and a mixture, we can consider a **spliced distribution** wherein we have a mixed variable, Y, defined as a mixture of component distributions, X_i, where each X_i is defined only on a certain interval, $[c_{i-1}, c_i)$. These intervals are always disjoint and are typically adjacent, and can be unionized to form a single piecewise fully defined interval. By piecewise fully defined we mean that when we take into account the continuity or discreteness of the component variable corresponding to the interval, $[c_{i-1}, c_i)$ every possible value between the lower bound and upper bound of the union will be accounted for. This gives the pdf of Y as $f_Y(y) = \alpha_i f_i(y)$ if $[c_{i-1}, c_i)$, where α_i is the mixing weight of X_i.

Example 2.5.1. *If X_1 is discretely uniform on $[0,5)$, X_2 is exponentially distributed with mean 8 on $[5,10)$, and X_3 has Poisson distribution with $\lambda = 20$ on $[10,25)$, and all $\alpha_i = \frac{1}{3}$, then Y has pdf*

$$f_Y(y) = \begin{cases} \left(\frac{1}{3}\right)\left(\frac{1}{5}\right) & if\ 0 \le y < 5 \\ \left(\frac{1}{3}\right)\left(\frac{1}{8}e^{-y/8}\right) & if\ 5 \le y < 10 \\ \left(\frac{1}{3}\right)\left(\frac{e^{-20}20^y}{y!}\right) & if\ 10 \le y < 25 \end{cases}$$

Another useful set of relations is when dealing with a conditional mixture is when the conditional distribution $Y|\Omega$, the conditioning distribution Ω, combine in a way that makes the unconditional distribution of Y into a distribution that we already know. For instance, if $Y|\Omega$ has exponential distribution with the mean Ω, and Ω has inverse gamma distribution with parameters α and θ, then the unconditional pdf of Y can be derived as follows.

$$\begin{aligned} f_Y(y) &= \int_0^\infty \left(\frac{1}{\lambda}e^{-y/\lambda}\right)\left(\frac{\theta^\alpha e^{-\theta/\lambda}}{\lambda^{\alpha+1}\Gamma(\alpha)}\right)d\lambda = \frac{\theta^\alpha}{\Gamma(\alpha)}\int_0^\infty \frac{e^{-(\theta+y)/\lambda}}{\lambda^{\alpha+2}}d\lambda \\ &= \frac{\theta^\alpha\Gamma(\alpha+1)}{(\theta+y)^{\alpha+1}\Gamma(\alpha)}\int_0^\infty \left(\frac{(\theta+y)^{\alpha+1}}{\Gamma(\alpha+1)}\right)\left(\frac{e^{-(\theta+y)/\lambda}}{\lambda^{\alpha+2}}\right)d\lambda \\ &= \frac{\theta^\alpha\Gamma(\alpha+1)}{(\theta+y)^{\alpha+1}\Gamma(\alpha)} \\ &= \frac{\alpha\theta^\alpha}{(\theta+y)^{\alpha+1}} \end{aligned}$$

This final result is actually a Pareto type 2 distribution with the same α and θ from the conditional inverse gamma distribution. Similarly, we can show a few other conditional distributions to have unconditional pdf belonging to known distributions. One is when $Y|\Omega$ has Poisson distribution with mean Ω, and Ω has gamma distribution with parameters α and θ, which results in an unconditional Y with negative binomial distribution with the parameters $r = \alpha$ and $\beta = \theta$. If $Y|\Omega$ has an inverse exponential distribution with parameter θ, Ω has gamma distribution with parameters α and θ, which yield an unconditional Y that has inverse Pareto distribution with parameters $\tau = \alpha$ and the same θ as that of Ω. Finally, if $Y|\Omega$ is normally distributed with a mean of Ω and variance of ρ^2, and Ω is also normally distributed with mean μ and variance σ^2 then the unconditional Y is another normally distributed variable with mean μ and variance $\rho^2 + \sigma^2$.

Chapter 3

Advanced Methods and Applications

In Chapter 2 we covered the basic methods used in short term actuarial modeling: using mixed distributions, maximum covered losses, and deductibles. These methods are useful on their own, but also act as a basis for some of the more advanced material. In this chapter, we will explore some of these advanced concepts that build on what we've already covered. Again, these methods are compiled in the Loss Models text and ACTEX manual, making them invaluable resources during research.

3.1 Aggregate Models

In application we are not usually concerned with values surrounding a single loss. Rather, we are concerned with the losses associated with a policy, which will have many clients covered under it. That is to say, there will be many people covered by a policy offered by an insurance provider and we want to analyze the performance and costs associated with *all* people who buy it. The individual losses, those associated with each client or customer, are represented by a **severity variable**. This variable is in terms of dollars and cents most of the time, and describes the *amount*, or monetary compensation, of insurance claims. It can be subject to deductibles or maximum covered losses, and can be either discrete or continuous in nature. The other variable we look at is called the **frequency variable**, which is representative of the *number* of claims under a given policy. Since it is impossible to have a fraction of a claim get filed, the frequency variable is always discrete and whole-numbered. With these, we can define the aggregate loss associated with an insurance policy.

Definition 3.1. *[Klu12] The **Aggregate Loss** is the sum of all losses associated with a policy in a given period of time. Aggregate losses are denoted as S and given by*

$$S = X_1 + X_2 + X_3 + ... + X_N$$

where all X_i are instances of the same severity variable and N is determined by the frequency variable.

When working with an aggregate loss model, we rely heavily on the conditional expected value and conditional variance as defined in the mixed variable section. This is because the expected value of S is completely determined by how many claims there are in a given time period and how much each loss is expected to be.

Theorem 3.2. *[Bro19] The expected total amount to be paid by an insurer on a policy with frequency N and severity X is,*

$$\begin{aligned} E[S] &= E[S|N] = E\big[(N)(E[X])\big] = E[N]E\big[E[X]\big] \\ &= E[N]E[X] \end{aligned}$$

As we see, the expected value of S works out very conveniently. The expected values of N and X will be given directly by some formula that we already know, either coming straight from the distribution's table or from a modification formula like that of an ordinary deductible applied to X. Similarly, we can work out a formula for the variance of S using the conditional variance formula.

Theorem 3.3. *[Bro19] The variance of the total amount paid by an insurer on a policy with frequency N and severity X is,*

$$\begin{aligned} Var[S] &= Var[E[S|N]] + E[Var[S|N]] = Var[(N)(E[X])] + E[(N)(Var[X])] \\ &= Var[N](E[X])^2 + E[N]Var[X] \end{aligned}$$

Example 3.1.1. *If X has Pareto Type 2 distribution with $\alpha = 3$ and $\theta = 20$ and N is Poisson with $\lambda = 15$, what is the expected value and variance of the aggregate variable S*

with frequency N and severity X?

$$E[S] = \left(\frac{20}{2}\right)(15) = \boldsymbol{150}$$
$$Var[S] = (15)\left(\frac{20}{2}\right)^2 + \left(\frac{2(20)^2}{(2)(1)} - \left(\frac{20}{2}\right)^2\right)(15) = \boldsymbol{6000}$$

Another important note to make is that when a modification is applied to the severity or frequency variable, we can still do all the same work and analysis associated with S by just using the modified versions of X or N when deriving values for S. Say X has a deductible applied to it, then $E[S] = E[N]E[(X-d)]$ and $Var[S] = Var[N](E[(X-d)])^2 + Var[(X-d)]E[N]$ where $E[(X-d)]$ and $Var[(X-d)]$ are found the same way that they were when analyzing $(X-d)$ on its own.

Example 3.1.2. *Consider the same distributions from Example 3.1.1, but where X has an ordinary deductible of* $d = 2$*, then*

$$E[S] = (15)(8.26) = \boldsymbol{123.90}$$
$$Var[S] = (15)(68.30) + (295.23)(15) = 1024.50 + 4428.45 = \boldsymbol{5452.95}$$

When we want to see likelihoods of values of S, it is usually best to use normal approximation, as writing the pdf and cdf of S can be a very long and arduous process. The normal approximation is accurate enough for our purposes, since this methodology is all predictive and won't be perfectly true-to-life. As such, even going through the complicated process of creating the actual pdf and cdf of S will not allow for perfect predictions, and can often times result in just as much, or even more, error than normal approximation just due to the amount of human work and rounding that must take place. After all, the typical case for an aggregate model will involve hundreds, if not thousands, of individual claims. For a brief review, normal approximation is when we define the variable $Z = \frac{S-E[S]}{\sqrt{Var[S]}}$. Z will always have a standard normal distribution, and will be reasonably accurate to probabilities of values for S. For probabilities of Z it is usually best to consult a normal distribution table, but the pdf of the normal distribution, $\phi(z) = (\frac{1}{\sqrt{2\pi}})e^{-z^2/2}$, works as well.

It is worth noting that we use N as our conditioning variable because the individual claim amounts will be independent of each other regardless of how many there are, but there is only a single N value which weighs heavily on S. Essentially, N is a stronger condition on S than X is, as it controls X in S.

3.2 Method of Convolution and Stop-Loss Insurance

Occasionally we may need to use the actual pdf and cdf of S rather than normal approximation. When this happens, we can actually write the functions of S in terms of those of X and N. This is done by using a rule of statistics that allows us to write, $P(X) = E[P(X|C)]$, where X is some event and C is some random variable. By using this, we can arrive at the following formulas for the pdf and cdf of S. Another method we must incorporate is the method of convolution, wherein we consider the number of individual losses, n from N, and express the probability of $S = s$ as the likelihood of n number of claims summing to be equal to s. This is called the ***nth-fold convolution of X***, which is denoted by X^{*n}. Applying convolution to discrete variables is a little different than applying it to continuous variables, so we must define the n-fold convolution for each case.

If X is discrete, then we consider convolution in a combinatorial sense. For instance, the probability of the 2-fold convolution being equal s will be written $P(X_1 + X_2 = s)$, and we must sum all probabilities of such pairs of X values. That is to say $P(X_1 + X_2 = s) = \sum_{j \epsilon \mathbf{Z}} P(X_1 = j \cap X_2 = s - j)$, where $\mathbf{Z}$ is the set of all integers. Since the individual X values will almost always be independent of each other, we can think of this as $\sum_{j \epsilon \mathbf{Z}} P(X_1 = j) \times P(X_2 = s - j)$. Furthermore, we can think of F_X^{*n} in generally the same way, going back to the 2-fold example this would be $P(X_1 + X_2 \leq s) = \sum_{j \epsilon \mathbf{Z}} P(X_1 = j) \times P(X_2 \leq s - j)$. For the general n-fold case, the work is the exact same, just repeated across n number of X_i instead of just two.

Theorem 3.4. *[Klu12] The nth-fold convolution of a discrete X will have the representative functions f_X^{*n} and F_X^{*n} which act as the pdf and cdf of X^{*n} respectively. These functions are,*

$$f_X^{*n}(s) = \sum_{j,k,\ldots,\alpha \epsilon \boldsymbol{Z}} P(X_1 = j) \times P(X_2 = k) \times \ldots \\ \times P(X_{n-1} = \alpha) \times P(X_n = s - (j + k + \ldots + \alpha))$$

$$F_X^{*n}(s) = \sum_{j,k,\ldots,\alpha \epsilon \boldsymbol{Z}} P(X_1 = j) \times P(X_2 = k) \times \ldots \\ \times P(X_{n-1} = \alpha) \times P(X_n \leq s - (j + k + \ldots + \alpha))$$

This may seem tedious in nature due to the need to repeatedly find and multiply probabilities in order to sum them, but in comparison to the continuous X case it is actually quite nice as long as we approach our combinations of X values in a systematic way. The continuous case works very much like the discrete case, but instead of repeatedly adding values we must integrate n times, once with respect to each X value.

Theorem 3.5. *[Klu12] The nth-fold convolution of a continuous X will have the representative functions f_X^{*n} and F_X^{*n} which act as the pdf and cdf of X^{*n} respectively. These functions are,*

$$f_X^{*n}(s) = \int_0^s \int_0^{s-j} ... \int_0^{s-(j+k+...+\alpha)} f_X(j) \times f_X(k) \times ... \times f_X(\alpha) \times f_X(s-(j+k+...+\alpha))d\alpha...dkdj$$

$$F_X^{*n}(s) = \int_0^s \int_0^{s-j} ... \int_0^{s-(j+k+...+\alpha)} f_X(j) \times f_X(k) \times ... \times f_X(\alpha) \times F_X(s-(j+k+...+\alpha))d\alpha...dkdj$$

Example 3.2.1. *Let S be a sum of three independent variables, X_1, X_2, and X_3. You are given the following information about these variables:*

$P(X_1 = 1) = 0.3$	$P(X_1 = 2) = 0.5$	$P(X_1 = 4) = 0.2$
$P(X_2 = 0) = 0.5$	$P(X_2 = 2) = 0.4$	$P(X_2 = 3) = 0.1$
$P(X_3 = 0) = 0.3$	$P(X_3 = 1) = 0.3$	$P(X_3 = 2) = 0.4$

Using convolution, we wish to find the probability that S=6. We first need all combinations of X_1, X_2, and X_3 available that add up to 6. Representing these as (X_1, X_2, X_3) we have (1, 3, 2); (2, 2, 2); (2, 3, 1); (4, 2, 0); and (4, 0, 2). Now we multiply the probabilities of the individual X_is in the ordered trios and add the products,

$$(0.3)(0.1)(0.4)+(0.5)(0.4)(0.4)+(0.5)(0.1)(0.3)+(0.2)(0.4)(0.3)+(0.2)(0.5)(0.4) = \boldsymbol{0.174}$$

So there is a 17.4% chance of S being 6.

Now that we have defined convolution for both the discrete and continuous severities, we can write the pdf and cdf of S using an n-fold convolution of X. We again assume that the X values are independent, as they usually will be in application since they represent insurance claims filed by different clients, and we also assume that the likelihood of $N = n$ is independent of the values that X takes. By doing this, we can write the pdf and cdf of S as follows.

$$F_S(s) = E[P(S \leq s|N)] = \sum_{n=0}^{\infty} F_X^{*n}(s)P(N = n) \tag{3.1}$$

$$f_S(s) = E[P(S = s|N)] = \sum_{n=0}^{\infty} f_X^{*n}(s)P(N = n) \tag{3.2}$$

These functions are not typically used when trying to find probabilities for values of S since normal approximation is faster, easier, and give approximate likelihoods that are accurate enough for use in application. Rather, the pdf and cdf of S will be used when considering the situation where S is under the effect of a deductible. We have previously analyzed the situation when an individual loss variable has a deductible applied to it, but another case to consider is when an aggregate model itself has a deductible. The notation will be the same, where $(S - d)$ represents S with an ordinary deductible of d. This is called the **Stop-Loss Insurance Payment**. Similar to the individual loss variable with a deductible, our primary concern will be with the expected value and variance of $(S - d)$, but there are other observations to be made. Also much like how the expected value of $(X - d)$ is formulated, we can use the pdf or cdf of S to find $E(S - d)$ using the same methods established previously.

[Bro19]
$$E[S-d] = \begin{cases} \int_d^{\infty}(y-d)f_S(y)dy & if\ S\ is\ continuous \\ \sum_{d+1}^{\infty}(k-d)f_S(k) & if\ S\ is\ discrete \\ \int_d^{\infty}[1-F_S(y)]dy & for\ either\ continuous\ or\ discrete\ S \end{cases} \tag{3.3}$$

Most of the time, the last case is best to use since it requires the least amount of work, just like when working with the individual loss variable. The term that is commonly used for $E[S - d]$ is the **net stop-loss premium**, but it is functionally the same as $E[X - d]$, just usually involving much larger amounts due to the nature of S.

Example 3.2.2. *Let S be a sum of two independent variables, X_1 and X_2, and subject to a deductible of 2. You are given the following information about X_1 and X_2:*

$P(X_1 = 1) = 0.3$	$P(X_1 = 2) = 0.4$	$P(X_1 = 3) = 0.3$
$P(X_2 = 1) = 0.1$	$P(X_2 = 3) = 0.2$	$P(X_2 = 4) = 0.7$

First let's list the possible values for S: {2, 3, 4, 5, 6, 7}. This tells us how many values we should be adding in our summation step, in this case we now know we will be adding six values.

$$\begin{aligned} E[S-2] = {} & (2-2)\big((0.3)(0.1)\big) + (3-2)\big((0.4)(0.1)\big) + (4-2)\big((0.3)(0.2) + (0.3)(0.1)\big) \\ & + (5-2)\big((0.3)(0.7) + (0.4)(0.2)\big) + (6-2)\big((0.4)(0.7) + (0.3)(0.2)\big) \\ & + (7-2)\big((0.3)(0.7)\big) \\ \\ & = 0 + 0.04 + 0.18 + 0.87 + 1.36 + 1.05 = 3.5 \approx 4 \end{aligned}$$

So the net top-loss premium is 4. We round to the nearest whole integer because both X_1 and X_2 being discrete forces their sum to be a whole number.

Finally, a useful observation is that since this works the same way as $E[X-d]$, we can apply the same property to $E[S-d]$ as we did to $E[X-d]$ in Theorem 2.7. Therefore,

$$E[S-d] = E[S] - E[S \wedge d] \tag{3.4}$$

3.3 Maximum Likelihood Estimation of Parametric Models

Sometimes we may encounter a set of data wherein we can see a general trend matching a known distribution, but we may not be able to easily tell what all of the parameters of the distribution are at first. In this situation, we can use **maximum likelihood estimation**, denoted mle, to find a functional value for the unknown parameter. There are two cases to consider for the mle, the first is when all values observed in the distribution are known exactly, and the other is when there are observed values that aren't exactly known. Values not known exactly are those that are less than the distribution's deductible, resulting in a 0 payment, or greater than its policy limit, resulting in a payment of u or $(u-d)$. The points below d or above u or $(u-d)$ are called **truncated data points**. In either case, we need to know some information about the distribution being worked with. We need to know what the distribution is, if there are multiple parameters then we need to know all but one, and we need to know some data points. As long as we know this information, we can develop a **Likelihood function** of parameter θ, denoted $L(\theta)$, based on the pdf of the distribution and the known data points. The idea is that once we have a likelihood function to work with, we can take its derivative with respect to θ, set that equal to zero and solve it for θ. This value, denoted $\widehat{\theta}$ will be our estimate, and will be accurate enough to work with.

3.3.1 MLE Based on Complete Data

When all observed values are known exactly we say that the data set is **complete**, in other words we must know that all values are non-truncated.

Theorem 3.6. *[Klu12] When all data points, $\{x_1, x_2, ..., x_n\}$, are known to be non-truncated, the* ***likelihood function*** *for θ is,*

$$L(\theta) = \prod_{i=1}^{n} f_X(x_i;\theta)$$

To simplify this equation, we may wish to consider the natural logarithm of $L(\theta)$. This will change the equation to allow us to use addition instead of multiplication, which will oftentimes make things much easier for us. If we have n data points, then the likelihood function of θ will involve a polynomial of at least degree n, but the degree could be any multiple of n depending on the distribution of X. For instance, if X has Pareto type 2 distribution, then the greatest degree term of the polynomial in $L(\theta)$ will be $\theta^{n(\alpha+1)}$ in the denominator of the product. Depending on the values for n and α in this example, we could easily end up with a very long and difficult to work through polynomial. But using $ln(L(\theta))$ will give us the same result, just typically with less work involved. Taking the natural log will result in the following equation, called the **Log-Likelihood function**.

Theorem 3.7. *[Klu12] When all data points, $\{x_1, x_2, ..., x_n\}$, are known to be non-truncated, the* ***log-likelihood function*** *for θ is,*

$$\ell(\theta) = ln\big(L(\theta)\big) = \sum_{i=1}^{n} ln\big(f(x_i, \theta)\big), \text{ with data points } x_1, x_2, ..., x_n$$

Example 3.3.1. *Suppose we have a variable X with Poisson distribution, unknown λ, and observed data points*

10, 10, 11, 13, 14, 15, 17

$$\ell(\lambda) = \sum_{i=1}^{7} ln\left(\frac{e^{-\lambda}\lambda^{k_i}}{k!}\right) = \sum_{i=1}^{7} -\lambda + k_i ln(\lambda) - ln(k_i!)$$
$$= -7\lambda + ln(\lambda) \times (2(10) + 11 + 13 + 14 + 15 + 17)$$
$$- ln(10! \times 10! \times 11! \times 13! \times 14! \times 15! \times 17!)$$
$$= -7\lambda + ln(\lambda) \times (90) - 156.86$$

This is the log-likelihood function we need for λ in our example. Now we set its derivative equal to zero and solve.

$$\widehat{\lambda} = -7 + \frac{90}{\lambda} = 0 \Rightarrow \frac{90}{\lambda} = 7 \Rightarrow \lambda = \frac{90}{7} = 12.857... \approx \boldsymbol{13}$$

In this case, we must round $\widehat{\lambda}$ to the nearest whole number, since X is a discrete Poisson variable with expected value equal to λ. Depending on the distribution, even some parameters of discrete distributions may not be whole numbers all the time. For instance, the parameter q in the binomial distribution has the property that $0 \leq q \leq 1$, so the only time that it will be a whole number is when the event X either never happens or is guaranteed.

3.3.2 MLE Based on Incomplete Data

If we know that the observed data points are subject to some modification, whether it's a deductible, a max covered loss, or both, and we have data points that are known to be truncated by these modifications, we say that the data set given is **incomplete** because not all data will be exactly known. When given an incomplete data set, we must consider three different cases: data is truncated only by a deductible, data is truncated only by a max covered loss, and data is truncated by both a max covered loss and a deductible. Recall as well that when discussing a deductible, we are really describing an ordinary deductible unless specified to be a franchise deductible.

If the data set only includes values truncated by a max covered loss, called **right-censored** data, then we must include in our likelihood function both the values that are below the mcl, u, and those that are equal to or above it. To do this we need to split the number of data points into two distinct tallies, where n represents the number of payments below u and m represents the number of payments above u. For those below u we can treat them as a complete subset of data, with a likelihood function for complete data. However, we must also include another function for the truncated data. This will be the survival function of each $(x_i \geq u)$, which is $\big(1 - F(u;\theta)\big)^m$. We multiply these two together in order to form the likelihood function for θ, because we assume that the individual data points are independent.

Theorem 3.8. *[Klu12] When the data set contains data points limited by a maximum covered loss, u, then the* ***likelihood function*** *of the parameter θ is,*

$$L(\theta) = \left(\prod_{i=1}^{n} f(x_i;\theta) \right) \times \big(1 - F(u;\theta)\big)^m$$

Taking the natural logarithm of this yields the ***log-likelihood function****,*

$$\ell(\theta) = \left(\sum_{i=1}^{n} ln\big(f(x_i;\theta)\big) \right) \times \Big(m \times ln\big(1 - F(u;\theta)\big)\Big)$$

Example 3.3.2. *Let X be a discrete variable that is uniformly distributed on [0, θ] and subject to a max covered loss of 70. We are given the following 10 data points:*

9, 11, 23, 34, 39, 45, 52, 67, 70, 70

So we can write the likelihood function,

$$L(\theta) = \Big(\prod_{i=1}^{8} \frac{1}{\theta} \Big) \times \Big(\prod_{j=1}^{2} \frac{\theta - 70}{\theta} \Big) = \Big(\frac{1}{\theta}\Big)^8 \times \Big(\frac{\theta - 70}{\theta}\Big)^2 = \frac{(\theta - 70)^2}{\theta^{10}}$$

Now we can take the natural log of L(θ), taking the derivative with respect to θ, and setting that equal to zero,

$$0 = \frac{2}{\theta - 70} - \frac{10}{\theta} = 2\theta - 10\theta + 700 \Rightarrow 8\theta = 700 \Rightarrow \theta = 87.5 \approx \boldsymbol{88}$$

So, $\widehat{\theta} = 88$

Something else to consider though is that the distribution of X could have mcls that are unique to each data point. This could occur when looking at insurance for clients of varying levels of risk; wherein the insurer may give a client of lower risk a higher mcl than a client of higher risk covered by the same general policy (i.e. all else remains the same). Under this circumstance, we can still derive a very similar likelihood and log-likelihood functions for θ.

$$L(\theta) = \left(\prod_{i=1}^{n} f(x_i;\theta)\right) \times \left(\prod_{j=1}^{m} (1 - F(u_j;\theta)\right) \tag{3.5}$$

[Bro19]

$$\ell(\theta) = \left(\sum_{i=1}^{n} ln\Big(f(x_i;\theta)\Big)\right) \times \left(\sum_{j=1}^{m} ln\Big(1 - F(u_j;\theta)\Big)\right) \tag{3.6}$$

Realistically, these formulations work just as well for the case where the same mcl is applied to all data points, since summing the same value m times is just m times that value; which is what we did in the case that all data points have the same mcl.

Now, we can also find the likelihood and log-likelihood functions for a parameter of a distribution that is being modified by a deductible based on a data set that includes value truncated by that deductible. These values, and by extension the data set, are called **left-censored**. There are two ways that these data sets can be given; the first is in the form $\{y_1, y_2, ..., y_n\}$ where y_i represents the actual payment made by the insurer, and the second is $\{x_1, x_2, ..., x_n\}$ where x_i represents the loss amounts before subtracting the deductible. These are logically equivalent since $y_i = x_i - d$, where d is the deductible being applied. Either way though, the probabilities are based on the distribution of the ground-up loss variable, X. So it will be the pdf and/or cdf of X that we use in the likelihood and log-likelihood functions. One difference to highlight is that the losses below the deductible would not likely be reported, so they won't usually be included in the data set. This is means that the number of data points below d will be unknown, so our approach will be different from that of the max covered loss case. Here, we must consider the data set to be conditional, meaning we think of these to be the observed values of X given that the observed value is greater than d.

The conditional nature of the data set leads to a conditional probability being used in the likelihood function of θ. Rather than using $f_X(x_i;\theta)$, we need to use $f_X(x_i;\theta|X > d)$. This yields the following likelihood and log-likelihood functions in terms of either x_i or y_i.

$$L(\theta) = \prod_{i=1}^{n} \frac{f_X(x_i;\theta)}{1 - F_X(d;\theta)} = \prod_{i=1}^{n} \frac{f_X(y_i + d;\theta)}{1 - F_X(d;\theta)} \tag{3.7}$$

[Bro19]

$$\ell(\theta) = \sum_{i=1}^{n} \Big(ln\big(f_X(x_i;\theta)\big) - ln\big(1 - F_X(d;\theta)\big) \Big) \tag{3.8}$$
$$= \sum_{i=1}^{n} \Big(ln\big(f_X(y_i + d;\theta)\big) - ln\big(1 - F_X(d;\theta)\big) \Big)$$

Example 3.3.3. *Consider an exponentially distributed variable X with a deductible of 15 and unknown parameter θ. We are given these 10 data points*

15, 19, 26, 30, 34, 40, 44, 46, 49, 53

and we want to find the mle of θ.

$$L(\theta) = \frac{\prod_{i=1}^{10} \frac{1}{\theta} e^{-x_i/\theta}}{(e^{-15/\theta})^{10}} = \Big(\frac{1}{\theta^{10}}\Big) e^{-(\Sigma x_i - 10(15))/\theta} = \Big(\frac{1}{\theta^{10}}\Big) e^{-(356-150)/\theta} = \Big(\frac{1}{\theta^{10}}\Big) e^{-216/\theta}$$
$$\Rightarrow \ell(\theta) = -10ln(\theta) - \frac{216}{\theta} \Rightarrow \ell(\theta)d\theta = -\frac{10}{\theta} + \frac{216}{\theta^2} = 0 \Rightarrow 10\theta = 216$$
$$\Rightarrow \theta = \boldsymbol{21.6}$$

So $\widehat{\theta} = 21.6$

Now, we can use these to consider the case that the values in the data set are not all subject to the same deductible. It works similar to the way that the different mcl's worked previously, but with one major difference from an understanding perspective. In the multiple mcl's case we had to consider the split data set where *some* values were modified by a u_i and the rest were not, but in this case *every* value will be modified by some d_i.

$$L(\theta) = \prod_{i=1}^{n} \frac{f_X(x_i;\theta)}{1 - F_X(d_i;\theta)} = \prod_{i=1}^{n} \frac{f_X(y_i + d_i;\theta)}{1 - F_X(d_i;\theta)} \tag{3.9}$$

[Bro19]

$$\ell(\theta) = \sum_{i=1}^{n} \Big(ln\big(f_X(x_i;\theta)\big) - ln\big(1 - F_X(d_i;\theta)\big) \Big) \tag{3.10}$$
$$= \sum_{i=1}^{n} \Big(ln\big(f_X(y_i + d_i;\theta)\big) - ln\big(1 - F_X(d_i;\theta)\big) \Big)$$

Something to note is that if data is given in terms of y_i then we may wish to use max likelihood estimation of a parameter of Y_P's distribution, rather than that of X. To do this, we treat the data set as complete and use $L(\theta) = \prod_{i=1}^{n} f_{Y_P}(y_i;\theta)$. This may not give an estimation for θ in the distribution of X, though. So if we need to estimate θ for the distribution of X then we need to identify the relationship between θ_{Y_P} and θ_X.

Finally, we may be given data that is under the effect of both an max covered loss and deductible. Any payment made in this case will be between 0 and $u - d$, where u is the mcl and d is the deductible. In terms of our data set, it is best to consider the data points to be y_i where $y_i = x_i - d$ and X is our ground-up loss variable. The likelihood functions here will look like hybrids of those of the deductible and max covered loss cases. We consider the n observed data points that aren't being right-censored, and m data points that are being right-censored; but the entire thing is conditional on there being $n + m$ data points that are greater than d. So, our likelihood and log-likelihood functions will be as follows.

$$L(\theta) = \frac{\Big(\prod_{i=1}^{n} f_X(x_i;\theta)\Big) \times \Big(1 - F_X(u;\theta)\Big)^m}{\Big(1 - F_X(d;\theta)\Big)^{n+m}} = \frac{\Big(\prod_{i=1}^{n} f_X(y_i + d;\theta)\Big) \times \Big(1 - F_X(u;\theta)\Big)^m}{\Big(1 - F_X(d;\theta)\Big)^{n+m}} \tag{3.11}$$

[Bro19]

$$\ell(\theta) = \sum_{i=1}^{n} \Big[ln\big(f_X(x_i;\theta)\big)\Big] + m\Big[ln\big(1 - F_X(u;\theta)\big)\Big] - (n+m)\Big[\big(1 - F_X(d;\theta)\big)\Big] \tag{3.12}$$

$$= \sum_{i=1}^{n} \Big[ln\big(f_X(y_i + d;\theta)\big)\Big] + m\Big[ln\big(1 - F_X(u;\theta)\big)\Big] - (n+m)\Big[\big(1 - F_X(d;\theta)\big)\Big] \tag{3.13}$$

These are usually quite sufficient when evaluating data sets that are both left and right-censored, but we can use the distribution of Y_P just like we did in the case of a left-censored set. When doing this, Y_P becomes the set that is *only* right-censored. The likelihood function of θ_{Y_P} will be $\big[\prod_{i=1}^{n} f_{Y_P}(y_i, \theta_{Y_P})\big] \times \big[1 - F_{Y_P}(u - d, \theta)\big]^m$. Observe also that if the data points are not all modified be the same deductible or mcl, the we can write the likelihood and log-likelihood functions for θ_X by just factoring out the corresponding components of the functions above and replacing them the same ways that we did when looking at the deductible and max covered loss cases on their own.

3.4 Special Cases of Aggregate Models and Maximum Likelihood Estimation

Occasionally, we may be met with specific conditions in our work that we can use to simplify the functions used when conducting our assessments of policies. For similar reasons to those in the previous chapter's overview of special cases for policy modifications, these conditions can result in significantly reduced functions being derived for the max likelihood estimation and aggregate loss model.

3.4.1 Aggregate Model Special Cases

The Poisson distribution is a fairly common discrete distribution that comes up often when looking at aggregate models. Predominantly it will come up as the distribution on N, the number of claims in a given period; in which case we say that S is a **compound Poisson** variable, regardless of the distribution of the individual losses, X. If S is compound Poisson, then we can drastically simplify the expected value and variance of S. Assuming λ to be the parameter of the Poisson distribution on N.

Theorem 3.9. *[Klu12] When S is an aggregate loss variable with severity X and frequency N, where N is Poisson with mean λ, then*

$$E[S] = E[N] \times E[X] = \lambda E[X]$$

$$\begin{aligned} Var[S] &= E[N] \times Var[X] + Var[N] \times (E[X])^2 \\ &= \lambda \times \left(E[X^2] - (E[X])^2\right) + \lambda \times (E[X])^2 = \lambda E[X^2] - \lambda(E[X])^2 + \lambda(E[X])^2 \\ &= \lambda E[X^2] \end{aligned}$$

The formula for $Var[S]$ is particularly useful if the distribution of X doesn't have a concise function for its variance. Were this the case, it may be quicker to work out the second moment of X and go back to the definition of variance, but using the above function we can skip the subtraction step of finding X's variance and just use the second moment that we were going to need anyway.

Example 3.4.1. *Consider the variable S that has severity variable X with Pareto Type 2 distribution and parameters $\alpha = 4$ and $\theta = 9$ and a frequency variable N with Poisson* *distribution and* $\lambda = 12$*. We wish to find* $E[S]$ *and* $Var[S]$*.*

$$E[S] = (12)\left(\frac{9}{3}\right) = (12)(3) = \boldsymbol{36}$$

$$Var[S] = (12)\left(\frac{2(9)^2}{(3)(2)}\right) = (12)(27) = \boldsymbol{324}$$

In this case, we would have needed to find both the first and second moments of X anyway, since the Pareto distribution doesn't have a function for variance. Using the special case's shortcut, we were able to skip the algebra involved in using the definition of variance for X.

Coincidentally, this is not the only nice way that Poisson can appear in an aggregate model. If $S_1, S_2, ..., S_m$ are all mutually independent compound Poisson variables with the Poisson parameters $\lambda_1, \lambda_2, ..., \lambda_m$, then $\Omega = \sum_{i=1}^{m} S_i$ will be a compound Poisson variable with the Poisson parameter $\lambda_\Omega = \sum_{i=1}^{m} \lambda_i$. It is important to make the distinction that we are dealing with compound Poisson variables and not standard Poisson variables. The key difference is that if the components are standard Poisson variables, then $E[S_i] = \lambda_i$, which would imply that $E[\Omega] = \lambda_\Omega \times E[M]$. This leads to drastically different values for $E[\Omega]$ and $Var[\Omega]$ than $E[S_i] = \lambda E[X_i^2]$. Another important distinction to make is that we aren't necessarily dealing with component compound Poisson variables with the same severity distribution, for instance S_1 could have a severity with Pareto Type 2 distribution and S_2 could have normally distributed losses.

We now have a λ_Ω for Ω's Poisson distribution, but what about the severity component? We must consider the severity to be a mixture of m variables. The variables being mixed will be the severity variables of the component compound Poisson variables, where the component severity X_i will have mixing weight $\frac{\lambda_i}{\lambda_\Omega}$. If X_i is the severity variable of S_i, we get the distribution function (cdf) of X_Ω to be $F_{X_\Omega}(x) = \sum_{i=1}^{m} \frac{\lambda_i}{\lambda_\Omega} \times F_i(x)$ and a pdf that is $f_{X_\Omega}(x) = \sum_{i=1}^{m} \frac{\lambda_i}{\lambda_\Omega} \times f_i(x)$ [Klu12]. From here we may be extra lucky and find that our severity mixture for Ω is actually a special mixture like one discussed in the previous chapter. Then we can more easily find the expected value and variance of X_Ω, making it easier to find the expected value and variance of Ω itself.

Example 3.4.2. *Let* S_1 *be compound Poisson with* $\lambda = 3$ *and* X_1 *being a continuously uniform variable on* $[0, 20]$*; and* S_2 *be compound Poisson with* $\lambda_2 = 5$ *and* X_2 *an exponentially distributed severity with* $\theta = 9$*. If* $\Omega = S_1 + S_2$*, find* $E[\Omega]$*.*

$$\lambda_\Omega = \lambda_1 + \lambda_2 = 3 + 5 = 8$$
$$E[X_\Omega] = \frac{3}{8}E[X_1] + \frac{5}{8}E[X_2] = \frac{3}{8}(10) + \frac{5}{8}(9) = \frac{30 + 45}{8} = \frac{75}{8}$$

So the expected value of Ω will be,

$$E[\Omega] = 8\left(\frac{75}{8}\right) = \boldsymbol{75}$$

3.4.2 Special Cases of Max Likelihood Estimation

When reading actuarial texts it is common to find several sections, if not several chapters, dedicated to the maximum likelihood functions of a number of commonly encountered distributions. While this is useful, most of these will not be made simpler by or highlight any special property of the distribution. Rather these chapters will be dedicated to just plugging distribution functions into the definitions of $L(\theta)$ and $\ell(\theta)$, for the sake of skipping some algebra later on down the line. This isn't to say that there aren't some interesting observations to make in regards to the mle's of certain distributions' parameters though.

The exponential distribution's maximum likelihood estimator for θ works out quite conveniently, thanks to the product of exponentials being sums of the exponents and the only variable increasing in degree being θ.

$$L(\theta) = \prod_{i=1}^{n} \frac{1}{\theta} e^{-x_i/\theta} = \frac{1}{\theta^n} e^{-(\sum x_i)/\theta} \Rightarrow \ell(\theta) = -n \times ln(\theta) - \frac{1}{\theta} \times \sum_{i=1}^{n} x_i$$
$$\Rightarrow \frac{d}{d\theta}\ell(\theta) = -\frac{n}{\theta} + \frac{1}{\theta^2} \times \sum_{i=1}^{n} x_i = 0 \Rightarrow \widehat{\theta} = \frac{\sum_{i=1}^{n} x_i}{n} \quad \text{[Bro19]}$$

This is simply the sample mean of X and is equal to the expected value of X, θ, which is very convenient. We can also use the distribution's lack of memory in a similar way to how we did in the previous chapter to find estimators for the incomplete data set possibilities. Since incomplete data results directly from modifications being placed on the ground-up loss variable, and the lack of memory property meaning that θ won't be impacted by the modifications, we can modify the sample mean slightly to properly estimate θ even when there are truncated data points in play. For instance, if the data set contains n non-truncated data points and m right-censored data points, then $\widehat{\theta} = \frac{1}{n} \times \left(\sum_{i=1} nx_i + m \times u \right)$. Similarly, if the data set is left-censored then $\widehat{\theta} = \frac{1}{n} \times \sum_{i=1}^{n} (x_i - d) = \frac{1}{n} \times \sum_{i=1}^{n} y_i$. Finally, if the data set contains n left-censored data points and m right-censored data points then $\widehat{\theta} = \frac{1}{n} \times \left(\sum_{i=1}^{n} (x_i - d) + m(u - d) \right)$. An interesting thing to notice is that for any exponential data set regardless of modifications being applied, $\widehat{\theta}$ is found by taking the total amount that that is being paid by the insurer and dividing that by the total number of non-right-censored data points.

Example 3.4.3. *Let X be an exponentially distributed loss variable with deductible d=5 and mcl u= 50. We are given the 10 data points,*

2, 10, 21, 30, 32, 37, 40, 43, 45, 45

where the payments of 45 are limit payments. Now we wish to find the max likelihood estimate of θ*.*

$$\widehat{\theta} = \frac{1}{10} \times \left(2 + 10 + 21 + 30 + 32 + 37 + 40 + 43 + 2(45)\right) = \frac{1}{10}(305)$$
$$= \boldsymbol{30.5}$$

Chapter 4

Application and Adaptation

The material covered in this project is predominantly geared towards the insurance industry, and seeks to predict behaviors of claims data based on policy decisions made. While this is the main way that these methods are used in the real world, they can also be powerful tools in other fields of analysis and prediction. Thus we will conclude with an example of application of some of the topics covered to an insurance policy, and two examples of how we may adapt what we have learned to anticipate outcomes in different scenarios that may not be so apparent.

4.1 MLE Applied to a Common Policy

Insurance policies tend to see changes in their claim numbers and amounts that maintain the original distribution, but with a slow change in the parameters involved. The following is an assessment of the possible change in parameter based on reported average claim amounts for a specific type of Medicare coverage for individuals suffering from exactly one chronic condition in 2011. At the time, the Medicare policy coverage had a median payout of $2722 [NLA11]. Note that we are not discussing the mean payout for this policy, this will be an important point of discussion later. The distribution of these claims is known to be Pareto Type 2, with parameter $\theta = 10$.

Say in the following year we observed these fifteen uncensored claim amounts in dollars: 1275; 1482; 1568; 1899; 1921; 2052; 2068; 2091; 2093; 2113; 2164; 2209; 2264; 2380; 2389. Assume we are able to tell through observation that θ is still 10. We now wish to find the new value for α in the claim distribution. Recall that the Pareto Type

2 distribution has the pdf

$$f_X(x) = \frac{\alpha\theta^\alpha}{(x+\theta)^{\alpha+1}}.$$

It is best to use the log-likelihood here, since the product of fifteen Pareto pdfs would be a monster of a polynomial to solve. So by applying Theorem 3.7 to construct the log-likelihood function yields,

$$\ell(\alpha) = \sum_{i=1}^{15} ln\left(\frac{\alpha(10)^\alpha}{(x_i+10)^{\alpha+1}}\right) = \sum_{i=1}^{15}\left[ln(\alpha) + \alpha ln(10) - (\alpha-1)ln(x_i+10)\right]$$

Notice that $ln(\alpha)$ and $\alpha ln(10)$ are constants that get repeatedly added to the function fifteen times, so we can pull those out of the summation and multiply them by 15. The $(\alpha-1)$ term also acts as a constant coefficient for the summation, so we can pull that out as well to narrow the scope of our recursive adding.

$$\ell(\alpha) = 15ln(\alpha) + 15ln(10) - (\alpha-1)\sum_{i=1}^{15} ln(x_i+10)$$

Since the summation will be long on its own, let's pull that aside and address it individually.

$$\sum_{i=1}^{15} ln(x_i+10) = ln(1285) + ln(1492) + ln(1578) + ln(1909) + ln(1931) + ln(2062)$$
$$+\, ln(2078) + ln(2101) + ln(2103) + ln(2123) + ln(2174) + ln(2219) + ln(2264)$$
$$+\, ln(2390) + ln(2399)$$
$$= \mathbf{75.194}$$

Now we plug this value into $\ell(\alpha)$ and distribute the $(\alpha-1)$ to continue.

$$\ell(\alpha) = 15ln(\alpha) + 15(\alpha)ln(10) - 75.194\alpha + 75.194$$

Now to solve for α we take the derivative of $\ell(\alpha)$ with respect to α and set it equal to 0.

$$0 = \frac{15}{\alpha} + 15ln(10) - 75.194 \Rightarrow 75.194 - 15ln(10) = \frac{15}{\alpha}$$
$$\Rightarrow \alpha = \frac{15}{40.655} = \mathbf{0.369}$$

So, we can think of our new Medicare payout distribution to be Pareto Type 2 with $\theta = 10$ and $\alpha = 0.369$. This is where the discussion becomes more interesting, because we notice that the expected value function for a Pareto Type 2 distribution relies on $\alpha > 1$. So does this mean that our α value is wrong? Not at all, in fact this is very normal in real-world application. Because the goal of these methods is to quantify social behaviors and things centered on human behavior, we may not always have values and parameters that work properly with our model, even if it is selected properly. This is why we often focus more on median value in application, rather than means.

The source of this issue is the necessity that the pdf of our variable being tracked must converge to some point in order for us to develop a working mean value. If the parameters are such that the function doesn't converge then no such mean value can be found; and since we are tracking human behavior, which is inconsistent, influenced heavily by emotion, and almost never perfectly fits any kind of logical model, we would expect to see many situations where our statistical models that are based on human behavior do not converge to some point. This doesn't mean that our model doesn't work when trying to make predictions about the behavior, just that there will be a higher margin for error. This is one way that some insurance providers determine when they need to start offering multiple new policies or policy classes [Klu12]. If there is consistent divergence that results in decreasingly accurate projections, then it is a sign that their members' needs no longer fit the policy as it was designed.

4.2 Adaptations of STAM Methods

Although the methods covered are already designed with insurance in mind, we can adapt them to work in different contexts in order to model other real-world scenarios. We will focus on two such scenarios here, first by using stop-loss insurance to model a potential bonus payment package and second by finding a likely interval for the damage caused by an earthquake hitting in a population center by using a conditional aggregate model.

4.2.1 Bonus Payment Modeling

A mining company uses expensive drilling equipment in order to tunnel through mountains, and although the machines are powerful they can break easily if improperly operated by employees. A supervisor notices that a specific equipment failure has occurred multiple times in the last six months, and it is costing the company a lot of money. The failure in question kills one of the main rubble conveyors completely, forcing the company to replace it for the full cost of $\$15,000$ [MIPR19]. To alleviate this, the supervisor hires a new technician to train the mining crews and offers him a bonus equal to the difference between $\$70,000$ and however much the company spends on repairing the same failure. The number of equipment failures follows a binomial distribution with parameters $q = 0.4$ and $m = 15$. The technician now wants to figure out how big of a bonus to expect, which we can do using similar methods to those used in stop-loss insurance.

We are given a hard ceiling of $70,000 for the technician's bonus, assuming he somehow brings the number of equipment replacements down to 0, so the bonus will be between $70,000 and $0. We must subtract from this ceiling the expected value of the cost associated with these machine replacements when under a maximum amount of $70,000. In notation, we write $E[Bonus] = 70,000 - E[S \wedge 70,000]$. This is very reminiscent of equation 3.4, where we said $E[S-d] = E[S]-E[S\wedge d]$, and will work out very similarly in calculation. We would usually apply the method of convolution now, but we can actually make things slightly easier for ourselves by noticing that since X is constant S must be discrete and that we can generate its pdf without resorting to convolution.

S is discrete, but its values are multiples of $\$15,000$. Furthermore, $P(S = w) = P(N = w/15,000)$. Since the pdf of N is $\binom{15}{k}(0.4)^k(0.6)^{15-k}$, we can substitute $w/15,000$ for k to get the pdf of S. So,

$$P(S = w) = \binom{15}{w/15,000}(0.4)^{w/15,000}(0.6)^{15-(w/15,000)}$$

From here we can treat this the same way we would if it were an individual random variable, and just go back to the original formula for $E[X\wedge u]$ to find $E[S\wedge 70,000]$, using Theorem 2.3.

$$
\begin{aligned}
E[S \wedge 70,000] &= \left(\sum_{w_i \leq 70,000} w_i \times P(S = w_i) \right) + 70,000\big(1 - P(S \leq 70,000)\big) \\
&= (60,000 \times 0.128) + (45,000 \times 0.0634) + (30,000 \times 0.0219) \\
&+ (15,000 \times 0.0049) + 70,000(1 - 0.2173) \\
&= 11,188.50 + 54,789 = \mathbf{65,977.50}
\end{aligned}
$$

Now that we have $E[S \wedge 70,000]$ we can find the expected bonus for the technician.

$$
E[Bonus] = 70,000 - 65,977.50 = \$\mathbf{4022.50}
$$

4.2.2 Earthquake Damage Modeling

Consider an earthquake striking in a major population center and causing a great deal of damage to buildings and structures. We know that both the damage caused per building and the number of buildings damaged will be dependent on distance from the epicenter of the earthquake, but they will not follow identical conditions. For instance, the damage caused per building will be strictly decreasing as we get further from the epicenter, as the energy is dispersing across a greater area. However, the number of buildings damaged may grow at first and then decay, since the number of buildings within the growing radius from the epicenter may grow faster than the quake's energy dissipates [Pat05]. This yields a double conditional aggregate model for damage caused by the quake.

Assume the severity of damage is represented by X and the number of buildings damaged is represented by N. X is exponentially distributed with mean Δ, a Pareto Type 2 variable with parameters $\theta_\Delta = 400$ and $\alpha_\Delta = 3$. N is a Poisson variable with mean Ω which is a gamma variable with $\alpha_\Omega = 30,000$ and $\theta_\Omega = 0.75$. We wish to find the interval given by $E[S] \pm \sigma_S$, where σ_S is the standard deviation of S. This interval is where we are most likely to find the actual damage caused after the event. Before going any further, it is good point out that N's conditional distribution is one of our special cases from Section 2.5.2. In this case, we can rewrite the distribution of N unconditionally as a negative binomial variable with parameters $\tau = \alpha_\Omega$ and $\beta = \theta_\Omega$.

Since both of our component variables are conditional, and we are working with an aggregate model, the conditional expected value and conditional variance functions from Theorem 2.15 and our aggregate expected value and aggregate variance functions from Theorems 3.2 and 3.3 should be applied in a few places here. First, we should find $E[S]$, which is actually pretty simple thanks to Theorem 3.2. $E[S] = E[N] \times E[X]$, but since N is a negative binomial variable $E[S] = \tau \times \beta = 0.75 \times 30,000 = 22,500$. Now we must apply Theorem 2.15 to X for its expected value, $E[X] = E\big[E[X|\Delta]\big] = E[\Delta]$ and $E[\Delta] = 400/(3-1) = 200$. Thus, $E[S] = 22,500 \times 200 = \mathbf{4,500,000}$. The standard deviation of S is a bit more involved though, since there are so many conditional components in it. Because of this, it is best to break it down into smaller parts. Recall that $Var[S]$ has its own formula, as outlined in Theorem 3.3.

$$Var[S] = Var[N] \times (E[X])^2 + Var[X] \times E[N].$$

In our case, all the components involving X must be calculated conditionally. Fortunately, we already found $E[X]$ so we only have $Var[X]$ left to solve for. We do have another conditional formula from Theorem 2.15 to find this though,

$$Var[X] = Var\big[E[X|\Delta]\big] + E\big[Var[X|\Delta]\big].$$

We already know that $E[X|\Delta] = \Delta$, and using the variance formula for the exponential distribution we know $Var[X|\Delta] = 2\Delta^2$. So we can now just plug these in and solve the equation using the distribution formulas for Δ.

$$Var[\Delta] = \frac{2(400)^2}{(2)(1)} - \left(\frac{400}{2}\right)^2 = 160,000 - 40,000 = 120,000$$
$$E[2\Delta^2] = 2E[\Delta^2] = 2\left(\frac{2(400)^2}{(2)(1)}\right) = 320,000$$

Therefore $Var[X] = 120,000 + 320,000 = \mathbf{440,000}$. Now we just apply the gamma distribution's variance formula to find $Var[N]$ and we will have the last piece of $Var[S]$.

$$Var[N] = 30,000 \times 0.75 \times 1.75 = 39,375$$

Bringing everything together for $Var[S]$ using Theorem 3.3 yields,

$$Var[S] = 39,375 \times 40,000 + 440,000 \times 22,500 = 11,475,000,000$$

Thus our standard deviation of S, which is the square root of $Var[S]$, is **107,121.43**. Now we can finally find the interval we set out to, $E[S] \pm \sigma_S$. The most likely damage range from the earthquake striking is **($4,392,878.57; $4,607,121.43)**.

Bibliography

[Bro19] Samuel A. Broverman. *ACTEX Exam STAM Study Manual, Fall 2019 Edition Volume 1*. ACTEX Learning, 2019.

[Pat05] Patricia Grisso, Howard Kunreuther, Chandu C. Patel. *Catastrophe Modeling: A New Approach To Managing Risk*. Springer Science+Business Media, 2005.

[Klu12] Stuart A. Klugman, Harry H. Panjer, Gordon E. Willmot. *Loss Models From: Data To Decisions, 4th Edition*. John Wiley & Sons, 2012.

[MIPR19] MIPR Corporation. *Heavy Industrial Belts*. Maine Industrial Corporation, 2019. *https://www.miprcorp.com/conveyor_belts/heavy-industrial-belts/*

[NLA11] National Library of Medicine. *Average Payment per Enrollee for Medicare Part A & B by Number of Chronic Conditions*. National Library of Medicine, 2011. *https://www.nlm.nih.gov/nichsr/stats_tutorial/section3/mod3_data.html*