Course/Exam 3 Study Note
Replacing the Chapter 2 Material from *Loss Models*[1]

Stuart A. Klugman

5th Printing, June 22, 2004[2]

[2] 2nd printing - In Example 5.6 (E[X]=1000, not 2000) and a correction to the solution for Exercise 19. 3rd printing - Definition 2.4 on page 6 has been changed which leads to changes in the following paragraph and Example 2.5. The use of loss elimination ratio is reduced and the definition of policy limit is clarified in pages 48–55. Example 5.15 on page 54 has been changed from 'per payment' to 'per loss' to make the answer correct. 4th printing - Several small changes have been made to improve clarity and readability. 5th printing - In Section 6.3.1 the correct count of 463 is used throughout.

Contents

Chapter 1

Introduction

The purpose of this Study Note is to replace some of the Exam[1] 3 material from *Loss Models*. When that book was written, the models/modeling split used in Courses 3 and 4 had not yet been established. As a result, the Exam 3 readings from this text are not contiguous, making it difficult for the reader to grasp the essential topics covered on Course 3. As well, the book was written to emphasize casualty losses and therefore did not convey the message that all actuarial models are essentially the same.

This material replaces the previously required sections[2] from Chapter 2 of *Loss Models*. Because the Chapter 3 material is more directly split (the beginning of each section appearing in Exam 3 and the end of each section in Course 4) that material will not be presented here. For those interested in another view of the material, the previously required sections from Chapter 2 are still relevant for Exam 3. However, any material that appears in that chapter but not in this Note will not be directly tested.

For this Note it is assumed that the reader has mastered basic probability concepts. Although many of them are reviewed here, the objective of this note is to concentrate on the particular probability concepts that are useful to actuaries and to introduce some relevant distributions that may not have been previously encountered.

Examples and exercises are scattered throughout. Exercises marked with (*) appeared in either a CAS 4B exam from 1996 through 1999 or in a Course 3 exam from May 2000 through May 2001. All of these questions were multiple choice, but the answer options are not given here. In addition, notation and terminology have been altered to match that used in this Note. Finally, although the nature of the material covered has been fairly constant, Exam 4B used the text *Loss Distributions* as source material and Exam 3 used *Loss Models* (which contains many 4B problems from 1995 and earlier). As always, caution should be used in extrapolating from past questions.

The topics to be covered in this Study Note are:

- Random variables as the common basis of actuarial models and some basic definitions and terminology.

[1] The Casualty Actuarial Society refers to its educational units as "exams" while the Society of Actuaries refers to its units as "courses." To save space while avoiding offending members of either group, the titles will be alternated throughout this note.

[2] For May 2001 the required sections were 2.1, 2.2 (Definitions 2.10, 2.11, 2.12, and 2.13 only), 2.6 (Pages 74–77 and 83 only), 2.7.1, 2.7.2 (excluding Example 2.40 and following), 2.7.3, 2.7.6, and 2.10 (excluding 2.10.1 and following).

- Introduction of several ongoing examples.
- Calculations using the models.
- Classifying models.
- Creating models.
- Additional calculations.
- Three examples.

A reasonable method for learning from this Note is to first read the text and examples from beginning to end. Then return to the beginning and work the exercises. Although this Note was reviewed extensively by a number of excellent people, it is possible that errors remain. Should you spot one I would appreciate it if you would inform me (e-mail me at Stuart.Klugman@Drake.edu). Also, if you have suggestions for improvements or areas that could use clarification, let me know.

Chapter 2

Random variables – the common basis of (most all) actuarial models

2.1 Introduction

The commonality we seek here is that all models for random phenomena have common elements. For each, there is a set of possible outcomes. The particular outcome that occurs will determine the success of our enterprise. Attaching probabilities to the various outcomes allows us to quantify our expectations and the risk of not meeting them. In this spirit, the underlying random variable will almost always be denoted X or Y. The context will provide a name and some likely characteristics. Of course, there are actuarial models that do not look like those covered on Course 3. For example, a "model office" is a list of cells containing policy type, age range, gender, etc.

To expand on this concept, consider the following definitions from the latest working draft of "Joint Principles of Actuarial Science."[1]

> *Phenomena* are occurrences that can be observed. An *experiment* is an observation of a given phenomenon under specified conditions. The result of an experiment is called an *outcome*; an *event* is a set of one or more possible outcomes. A stochastic phenomenon is a phenomenon for which an associated experiment has more than one possible outcome. An event associated with a *stochastic phenomenon* is said to be *contingent*. *Probability* is a measure of the likelihood of the occurrence of an event. It is measured on a scale of increasing likelihood from zero to one. A *random variable* is a function that assigns a numerical value to every possible outcome.

The following list contains a number of random variables encountered in actuarial work.

1. The age at death of a randomly selected birth. **(Model 1)**

2. The time to death from when insurance was purchased for a randomly selected insured life.

[1] This document is a work in progress of a joint committee from the Casualty Actuarial Society and the Society of Actuaries. Key principles are that models exist that represent actuarial phenomena (Exam 3) and that given sufficient data it is possible to calibrate models (Course 4).

3. The time from occurrence of a disabling event to recovery or death for a randomly selected workers compensation claimant.

4. The time from the incidence of a randomly selected claim to its being reported to the insurer.

5. The time from the reporting of a randomly selected claim to its settlement.

6. The number of dollars paid on a randomly selected life insurance claim.

7. The number of dollars paid on a randomly selected automobile bodily injury claim. **(Model 2)**

8. The number of automobile bodily injury claims in one year from a randomly selected insured automobile. **(Model 3)**

9. The total dollars in medical malpractice claims paid in one year owing to events at a randomly selected hospital. (**Model 4**)

10. The time to default or prepayment on a randomly selected insured home loan that terminates early.

11. The amount of money paid at maturity on a randomly selected high-yield bond.

Because all of these phenomena can be expressed as random variables, the machinery of probability and mathematical statistics is at our disposal both to create and to analyze models for them. The following paragraphs discuss the five key functions used in describing a random variable. They will be illustrated with four ongoing models as identified in the list above plus two more to be introduced later.

2.2 Key functions and four models

Definition 2.1 *The* ***cumulative distribution function****, also called the* ***distribution function*** *and usually denoted* $F_X(x)$ *or* $F(x)$*,*[2] *for a random variable* X *is the probability that* X *is less than or equal to a given number. That is,* $F_X(x) = \Pr(X \le x)$*. The abbreviation* ***cdf*** *is often used.*

The distribution function must satisfy a number of requirements:[3]

- $0 \le F(x) \le 1$ for all x.
- $F(x)$ is non-decreasing.
- $F(x)$ is right-continuous.[4]

[2] When denoting functions associated with random variables, it is common to identify the random variable through a subscript on the function. In this note, subscripts will be used only when needed to distinguish one random variable from another. In addition, for the five models to be introduced shortly, rather than write the distribution function for random variable 2 as $F_{X_2}(x)$, it will simply be denoted $F_2(x)$.

[3] The first item is not necessary, it follows from the next three.

[4] Right-continuous means that at any point x_0, the limiting value of $F(x)$ as x approaches x_0 from the right is equal to $F(x_0)$. This need not be true as x approaches x_0 from the left.

- $\lim_{x\to-\infty} F(x) = 0$ and $\lim_{x\to\infty} F(x) = 1$.

Because it need not be left-continuous, it is possible for the distribution function to jump. When it jumps, the value is assigned to the top of the jump.

Here are possible distribution functions for each of the four models.

Model 1[5] *This random variable could serve as a model for the age at death. It is continuous and cannot take on values beyond 100. (An informal definition of a continuous distribution is that over the range where the distribution function is not zero or one, it is differentiable. An interpretation is that all real values in that range are possible.) While experience suggests that there is an upper bound for human lifetime, models with no upper limit may be useful if they assign extremely low probabilities to extreme ages. This allows the modeler to avoid setting a specific maximum age.*

$$F_1(x) = \begin{cases} 0, & x < 0 \\ 0.01x, & 0 \le x < 100 \\ 1, & x \ge 100. \end{cases}$$

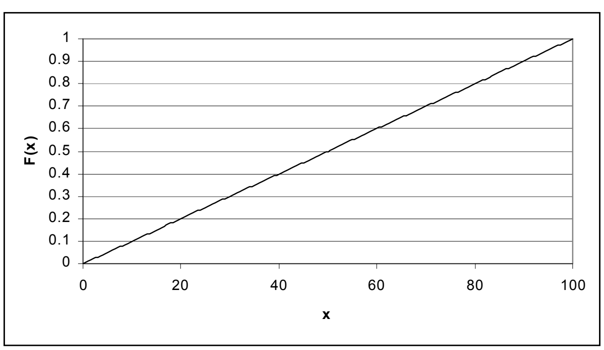
□

Model 2 *This random variable could serve as a model for the number of dollars paid on an automobile insurance claim. It is also continuous, but is unbounded. As with mortality, there is more than likely an upper limit (all the money in the world comes to mind), but this model illustrates that in modeling, correspondence to reality need not be perfect.*

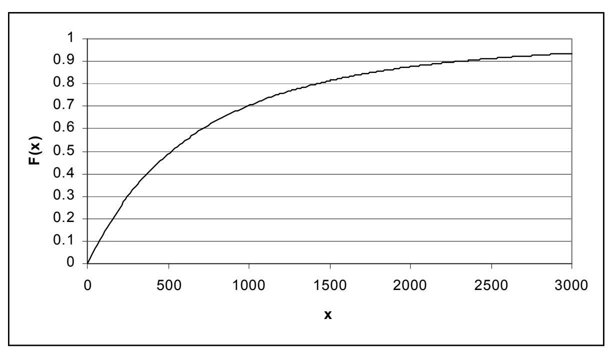
$$F_2(x) = \begin{cases} 0, & x < 0 \\ 1 - \left(\frac{2000}{x+2000}\right)^3, & x \ge 0. \end{cases}$$

□

Example 2.2 *Draw graphs of the distribution function for Models 1 and 2 (graphs for the other models are requested in Exercise 2).*



Distribution Function for Model 1



Distribution Function for Model 2

□

Model 3 *This random variable could serve as a model for the number of claims on one policy in one year. It is discrete, meaning that probability is concentrated at a countable number of points.*

[5]The six models (four introduced here and two later) will be identified by the numbers 1–6. Other examples will use the traditional numbering scheme as used for Definitions, etc.

Here there are only five (0,1,2,3,4) such points and the probability at each is given by the size of the jump in the distribution function. Discrete models can take on a finite number of possible values (as in this model) or a countably infinite number of values (such as with the Poisson distribution).

$$F_3(x) = \begin{cases} 0, & x < 0 \\ 0.5, & 0 \le x < 1 \\ 0.75, & 1 \le x < 2 \\ 0.87, & 2 \le x < 3 \\ 0.95, & 3 \le x < 4 \\ 1, & x \ge 4. \end{cases}$$

□

Model 4 *This random variable could serve as a model for the total dollars paid on a malpractice policy in one year. It is a mixed distribution. There is probability 0.7 that no dollars will be paid (so this probability is discrete) and the remaining 0.3 of probability is distributed continuously over positive values.*

$$F_4(x) = \begin{cases} 0, & x < 0 \\ 1 - 0.3e^{-0.00001x}, & x \ge 0. \end{cases}$$

□

Definition 2.3 *The **support** of a random variable is the set of numbers that are possible values of the random variable.*

Definition 2.4 *A random variable is called **discrete** if the support contains at most a countable number of values. It is called **continuous** if the distribution function is continuous and is differentiable everywhere with the possible exception of a countable number of values. It is called **mixed** if it is not discrete and is continuous everywhere with the exception of at least one value and at most a countable number of values.*

These three definitions do not exhaust all possible random variables, but will cover all cases encountered in this Note. The distribution function for a discrete variable will be constant except for jumps at the values with positive probability. A mixed distribution will have at least one jump. Requiring continuous variables to be differentiable allows the variable to have a density function (defined later) at almost all values.

Example 2.5 *For each of the four models, determine the support and indicate which type of random variable it is.*

The distribution function for Model 1 is continuous and is differentiable except at 0 and 100 and therefore is a continuous distribution. The support is values from 0 to 100 with it not being clear if 0 or 100 are included. The distribution function for Model 2 is continuous and is differentiable except at 0 and therefore is a continuous distribution. The support is all positive numbers and perhaps 0. The random variable for Model 3 places probability only at 0, 1, 2, 3, and 4 (the support) and thus is discrete. The distribution function for Model 4 is continuous except at 0 where it jumps. It is a mixed distribution with support on non-negative numbers. □

These four models illustrate the most commonly encountered forms of the distribution function. Graphs of these distribution functions (as well as other functions to be introduced in this section) are requested in Exercise 2 on Page 11 and the graphs themselves appear in the Appendix. For the remainder of this note, values of functions like the distribution function will be presented only for values in the range of the support of the random variable.

Definition 2.6 *The **survival function**, usually denoted $S_X(x)$ or $S(x)$, for a random variable X is the probability that X is greater than a given number. That is, $S_X(x) = \Pr(X > x) = 1 - F_X(x)$.*

As a result,

- $0 \le S(x) \le 1$ for all x.
- $S(x)$ is non-increasing.
- $S(x)$ is right-continuous.
- $\lim_{x\to-\infty} S(x) = 1$ and $\lim_{x\to\infty} S(x) = 0$.

Because the survival function need not be left-continuous, it is possible for it to jump (down). When it jumps, the value is assigned to the bottom of the jump.

Because the survival function is the complement of the distribution function, knowledge of one implies knowledge of the other. Historically, when the random variable is measuring time, the survival function is presented, while when it is measuring dollars, the distribution function is presented.

Example 2.7 *For completeness, here are the survival functions for the four models.*

$$S_1(x) = 1 - 0.01x,\ 0 \le x < 100$$

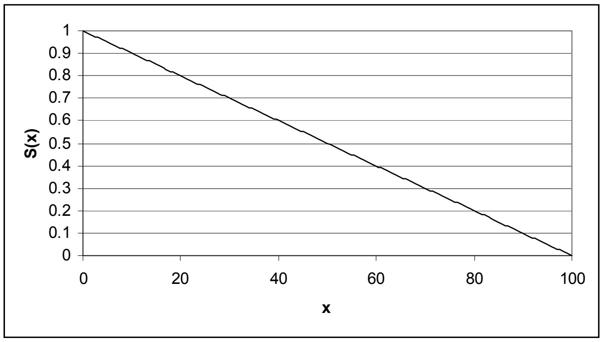
$$S_2(x) = \left(\frac{2000}{x+2000}\right)^3,\ x \ge 0$$

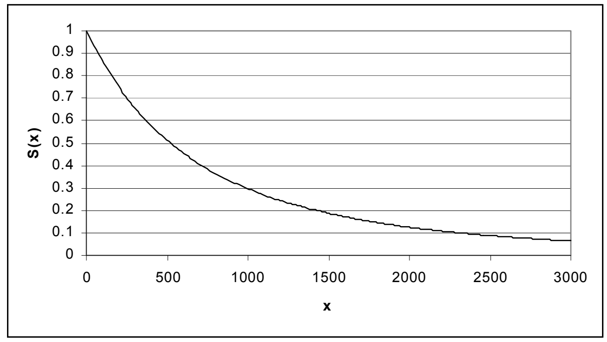
$$S_3(x) = \begin{cases} 0.5, & 0 \le x < 1 \\ 0.25, & 1 \le x < 2 \\ 0.13, & 2 \le x < 3 \\ 0.05, & 3 \le x < 4 \\ 0, & x \ge 4 \end{cases}$$

$$S_4(x) = 0.3e^{-0.00001x},\ x \ge 0.$$

□

Example 2.8 *Graph the survival function for Models 1 and 2.*





□

Survival Function for Model 1 Survival Function for Model 2

Either the distribution or survival function can be used to determine probabilities. Let $F(b-) = \lim_{x \nearrow b} F(x)$ and let $S(b-)$ be similarly defined. That is, we want the limit as x approaches b from below. We have $\Pr(a < X \leq b) = F(b) - F(a) = S(a) - S(b)$ and $\Pr(X = b) = F(b) - F(b-) = S(b-) - S(b)$. When the distribution function is continuous at x, $\Pr(X = x) = 0$; otherwise the probability is the size of the jump. The next two functions are more directly related to the probabilities. The first is for continuous distributions, the second for discrete distributions.

Definition 2.9 *The* ***probability density function****, also called the* ***density function****, usually denoted $f_X(x)$ or $f(x)$, is the derivative of the distribution function, or equivalently, the negative of the derivative of the survival function. That is, $f(x) = F'(x) = -S'(x)$. The density function is defined only at those points where the derivative exists. The abbreviation* ***pdf*** *is often used.*

While the density function does not directly provide probabilities, it does provide relevant information. Values of the random variable in regions with higher density values are more likely to occur than those in regions with lower values. Probabilities for intervals and the distribution and survival functions can be recovered by integration. That is, when the density function is defined over the relevant interval, $\Pr(a < X \leq b) = \int_a^b f(x)dx$, $F(b) = \int_{-\infty}^b f(x)dx$, and $S(b) = \int_b^\infty f(x)dx$.

Example 2.10 *For our models, we have*

$$f_1(x) = 0.01, \quad 0 < x < 100$$

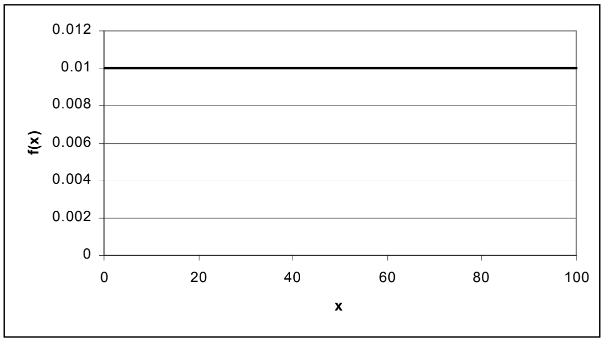
$$f_2(x) = \frac{3(2000)^3}{(x+2000)^4}, \quad x > 0$$

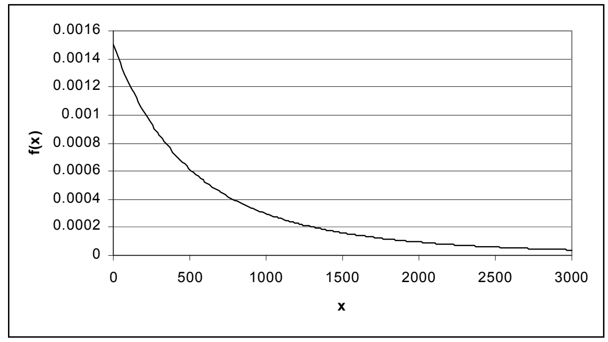
$$f_3(x) \text{ is not defined}$$

$$f_4(x) = 0.000003e^{-0.00001x}, \quad x > 0.$$

It should be noted that for Model 4 the density function does not completely describe the probability distribution. As a mixed distribution, there is also discrete probability at 0. □

Example 2.11 *Graph the density function for models 1 and 2.*





□

Density Function for Model 1 Density Function for Model 2

Definition 2.12 *The* ***probability function,*** *also called the* ***probability mass function****, usually denoted $p_X(x)$ or $p(x)$, describes the probability at a distinct point when it is not 0. The formal definition is $p_X(x) = \Pr(X = x)$.*

For discrete random variables, the distribution and survival functions can be recovered as $F(x) = \sum_{y \leq x} p(y)$ and $S(x) = \sum_{y > x} p(y)$.

Example 2.13 *The four models become*

$$p_1(x) \text{ is not defined}$$

$$p_2(x) \text{ is not defined}$$

$$p_3(x) = \begin{cases} 0.50, & x = 0 \\ 0.25, & x = 1 \\ 0.12, & x = 2 \\ 0.08, & x = 3 \\ 0.05, & x = 4 \end{cases}$$

$$p_4(0) = 0.7.$$

It is again noted that the distribution in Model 4 is mixed, so the above describes only the discrete portion of that distribution. There is no easy way to present probabilities/densities for a mixed distribution. On the other hand, they tend to be more revealing of the mixed nature of the distribution. For Model 4 we would present the probability density function as

$$f_4(x) = \begin{cases} 0.7, & x = 0 \\ 0.000003e^{-0.00001x}, & x > 0 \end{cases}$$

realizing that, technically, it is not a probability density function at all. When the density function is assigned a value at a specific point, as opposed to being defined on an interval, it is understood to be a discrete probability mass. □

Definition 2.14 *The **hazard rate**, also known as the **force of mortality** and the **failure rate** and usually denoted $h_X(x)$ or $h(x)$, is the ratio of the density and survival functions, when the density function is defined. That is, $h_X(x) = f_X(x)/S_X(x)$.*

When called the force of mortality, the hazard rate is often denoted $\mu(x)$ and when called the failure rate it is often denoted $\lambda(x)$. Regardless, it may be interpreted as the probability density at x given that the argument will be at least x. We also have $h_X(x) = -S'(x)/S(x) = -d\ln S(x)/dx$. The survival function can be recovered from $S(b) = e^{-\int_{-\infty}^{b} h(x)dx}$. In mortality terms, it is the annualized probability[6] that a person age x will die in the next instant, expressed as a death rate per year. In this note we will always use $h(x)$ to denote the hazard rate, although one of the alternative names may be used.

Example 2.15 *For our models, we have*

$$h_1(x) = \frac{0.01}{1-0.01x}, \quad 0 < x < 100$$

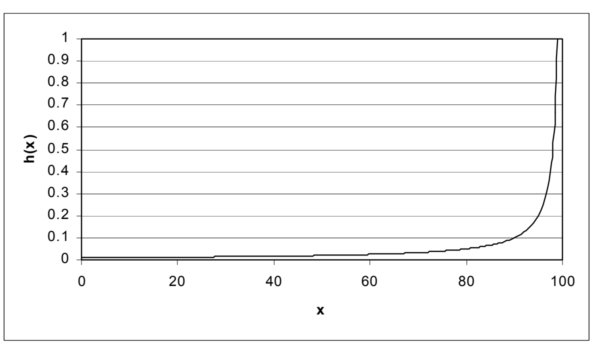
$$h_2(x) = \frac{3}{x+2000}, \quad x > 0$$

$$h_3(x) \text{ is not defined}$$

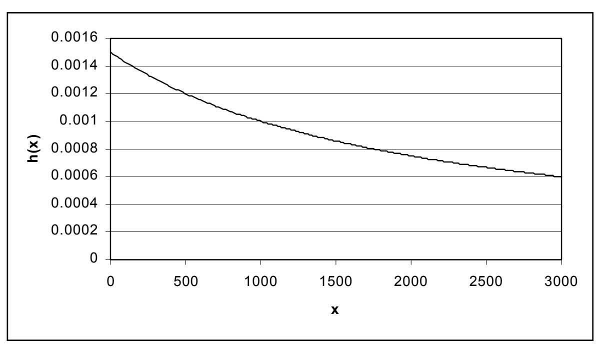
$$h_4(x) = 0.00001, \quad x > 0.$$

Once again, note that for the mixed distribution, the hazard rate is only defined over part of the random variable's support. This is different from the problem above where both a probability density function and a probability function are involved. Where there is a discrete probability mass, the hazard rate is not defined. □

Example 2.16 *Graph the hazard rate function for Models 1 and 2.*



Hazard Rate Function for Model 1



Hazard Rate Function for Model 2

□

[6] Note that the force of mortality is not a probability (in particular, it can be greater than 1) although it does no harm to visualize it as a probability.

The following model illustrates a situation in which there is a point where the density and hazard rate functions are not defined.

Model 5 *An alternative to the simple lifetime distribution in Model 1 is given below. Note that it is piecewise linear and the derivative at 50 is not defined. Therefore, neither the density function nor the hazard rate function is defined at 50. Unlike the mixed model of Model 4, there is no discrete probability mass at this point. Because the probability of 50 occurring is zero, the density or hazard rate at 50 could be arbitrarily defined with no effect on subsequent calculations. In this note such values will be arbitrarily defined so that the function is right continuous.*[7] *See the solution to Exercise 1 for an example.*

$$S_5(x) = \begin{cases} 1 - 0.01x, & 0 \leq x < 50 \\ 1.5 - 0.02x, & 50 \leq x < 75. \end{cases}$$

□

Exercise 1 *Determine the distribution, density, and hazard rate functions for Model 5.*

Appendix A from *Loss Models* presents the density and distribution functions for a number of commonly used continuous random variables. It will be referred to as *LMA* in this note.

Exercise 2 *Construct graphs of the distribution function for Models 3–5. Also graph the density or probability function as appropriate and the hazard rate function, where it exists.*

An interesting feature of a random variable is the value that is most likely to occur.

Definition 2.17 *The* **mode** *of a random variable is the most likely value. For a discrete variable it is the value with the largest probability. For a continuous variable it is the value for which the density function is largest. If there are local maxima, these points are also considered to be modes.*

Example 2.18 *Where possible, determine the mode for Models 1–5.*

For Model 1, the density function is constant. All values from 0 to 100 could be the mode, or equivalently, it could be said that there is no mode. For Model 2, the density function is strictly decreasing and so the mode is at 0. For Model 3, the probability is highest at 0. As a mixed distribution, it is not possible to define a mode for Model 4. Model 5 has a density that is constant over two intervals, with higher values from 50 to 75. These values are all modes. □

Exercise 3 *(*) A random variable X has density function $f(x) = 4x(1+x^2)^{-3}$, $x > 0$. Determine the mode of X.*

Exercise 4 *(*) A non-negative random variable has a hazard rate function of $h(x) = A+e^{2x}$, $x \geq 0$. You are also given $S(0.4) = 0.5$. Determine the value of A.*

Exercise 5 *(*) X has a Pareto distribution with parameters $\alpha = 2$ and $\theta = 10{,}000$. Y has a Burr distribution with parameters $\alpha = 2$, $\gamma = 2$, and $\theta = \sqrt{20{,}000}$. Let r be the ratio of $\Pr(X > d)$ to $\Pr(Y > d)$. Determine $\lim_{d\to\infty} r$.*

[7] By arbitrarily defining the value of the density or hazard rate function at such points, it is clear that using either of them to obtain the survival function will work. If there is discrete probability at this point (in which case these functions are left undefined) then the density and hazard functions are not sufficient to completely describe the probability distribution.

Chapter 3

Basic calculations using distributional models

3.1 Moments

There are a variety of interesting calculations that can be done from the models described in the previous chapter. Examples are the average amount paid on a claim that is subject to a deductible or policy limit or the average remaining lifetime of a person age 40.

Definition 3.1 *The* ***kth raw moment*** *of a random variable is the expected (average) value of the kth power of the variable, provided it exists. It is denoted by $E(X^k)$ or by μ'_k. The first raw moment is called the* ***mean*** *of the random variable and is usually denoted by μ.*

Note that μ is not related to $\mu(x)$, the force of mortality as mentioned on Page 10. For random variables that take on only positive values (that is, $\Pr(X > 0) = 1$), k may be any real number. When presenting formulas for calculating this quantity, a distinction between continuous and discrete variables needs to be made. Formulas will be presented for random variables that are either everywhere continuous or everywhere discrete. For mixed models, evaluate the formula by integrating with respect to its density function wherever the random variable is continuous and by summing with respect to its probability function wherever the random variable is discrete and adding the results. The formula for the kth raw moment is

$$\begin{aligned} \mu'_k = E(X^k) &= \int_{-\infty}^{\infty} x^k f(x)dx, \text{ if the random variable is continuous} \\ &= \sum_i x_i^k p(x_i), \text{ if the random variable is discrete} \end{aligned} \tag{3.1}$$

where the sum is to be taken over all x_i with positive probability. Finally, note that it is possible that the integral or sum will not converge, in which case the moment is said not to exist.

Example 3.2 *Determine the first two raw moments for each of the five models.*

The subscripts on the random variable X indicate which model is being used.

$$E(X_1) = \int_0^{100} x(0.01)dx = 50, \qquad E(X_1^2) = \int_0^{100} x^2(0.01)dx = 3{,}333.33$$

$$E(X_2) = \int_0^{\infty} x\frac{3(2000)^3}{(x+2000)^4}dx = 1{,}000, \qquad E(X_2^2) = \int_0^{\infty} x^2\frac{3(2000)^3}{(x+2000)^4}dx = 4{,}000{,}000$$

$$E(X_3) = 0(.5) + 1(.25) + 2(.12) + 3(.08) + 4(.05) = 0.93$$

$$E(X_3^2) = 0(.5) + 1(.25) + 4(.12) + 9(.08) + 16(.05) = 2.25$$

$$E(X_4) = 0(.7) + \int_0^{\infty} x(.000003)e^{-0.00001x}dx = 30,000$$

$$E(X_4^2) = 0^2(.7) + \int_0^{\infty} x^2(.000003)e^{-0.00001x}dx = 6{,}000{,}000{,}000$$

$$E(X_5) = \int_0^{50} x(.01)dx + \int_{50}^{75} x(.02)dx = 43.75$$

$$E(X_5^2) = \int_0^{50} x^2(.01)dx + \int_{50}^{75} x^2(.02)dx = 2{,}395.83.$$

□

Before proceeding further, an additional model will be introduced. This one looks similar to Model 3, but with one key difference. It is discrete, but with the added requirement that all of the probabilities must be integral multiples of some number. The Course 3/4 split treats the models in Exam 3 as given (or perhaps as known from some prior investigations). It is not until Course 4 that we investigate the source of the models and learn how to use data to construct them. However, there is one model that is so closely tied to the data that we cannot wait until Exam 4 to introduce it.

Definition 3.3 *The **empirical model** is a discrete distribution based on a sample of size n which assigns probability $1/n$ to each data point.*

Model 6 *Consider a sample of size 8 in which the observed data points were 3, 5, 6, 6, 6, 7, 7, and 10. The empirical model then has probability function*

$$p_6(x) = \begin{cases} 0.125, & x = 3 \\ 0.125, & x = 5 \\ 0.375, & x = 6 \\ 0.25, & x = 7 \\ 0.125, & x = 10. \end{cases}$$

□

Alert readers will note that many discrete models with finite support look like empirical models. Model 3 could have been the empirical model for a sample of size 100 that contained 50 zeros, 25 ones, 12 twos, 8 threes, and 5 fours. Regardless, we will use the term empirical model only when there is an actual sample behind it. The two moments for Model 6 are

$$E(X_6) = 6.25 \quad E(X_6^2) = 42.5$$

using the same approach as in Model 3. It should be noted that the mean of this random variable is equal to the sample arithmetic average (also called the sample mean).

Definition 3.4 *The* ***kth central moment*** *of a random variable is the expected value of the* k*th power of the deviation of the variable from its mean. It is denoted by* $E[(X-\mu)^k]$ *or by* μ_k*. The second central moment is usually called the* ***variance*** *and denoted* σ^2 *and its square root,* σ*, is called the* ***standard deviation****. The ratio of the standard deviation to the mean is called the* ***coefficient of variation****. The ratio of the third central moment to the cube of the standard deviation,* $\gamma_1 = \mu_3/\sigma^3$*, is called the* ***skewness****. The ratio of the fourth central moment to the fourth power of the standard deviation,* $\gamma_2 = \mu_4/\sigma^4$*, is called the* ***kurtosis****.*[1]

The continuous and discrete formulas for calculating central moments are

$$\begin{aligned}\mu_k = E[(X-\mu)^k] &= \int_{-\infty}^{\infty}(x-\mu)^k f(x)dx, \text{ if the random variable is continuous} \\ &= \sum_i (x_i-\mu)^k p(x_i), \text{ if the random variable is discrete.}\end{aligned} \tag{3.2}$$

In reality, the integral need be taken only over those x values where $f(x)$ is positive. The standard deviation is a measure of how much the probability is spread out over the random variable's possible values. It is measured in the same units as the random variable itself. The coefficient of variation measures the spread relative to the mean. The skewness is a measure of asymmetry. A symmetric distribution has a skewness of zero, while a positive skewness indicates that probabilities to the right tend to be assigned to values further from the mean than those to the left. The kurtosis measures flatness of the distribution relative to a normal distribution (which has a kurtosis of 3). Kurtosis values above 3 indicate that (keeping the standard deviation constant) relative to a normal distribution, more probability tends to be at points away from the mean than at points near the mean. The coefficient of variation, skewness, and kurtosis are all dimensionless.

There is a link between raw and central moments. The following equation indicates the connection between second moments. The development uses the continuous version from Equations (3.1) and (3.2) but the result applies to all random variables.

$$\begin{aligned}\mu_2 &= \int_{-\infty}^{\infty}(x-\mu)^2 f(x)dx = \int_{-\infty}^{\infty}(x^2-2x\mu+\mu^2)f(x)dx \\ &= E(X^2) - 2\mu E(X) + \mu^2 = \mu_2' - \mu^2.\end{aligned} \tag{3.3}$$

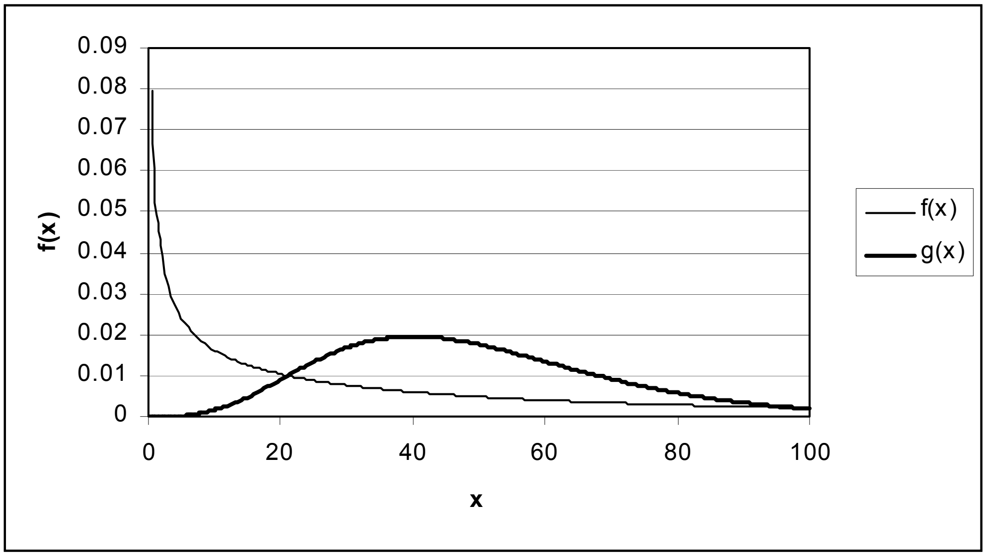
Exercise 6 *Develop formulas similar to Equation (3.3) for* μ_3 *and* μ_4*.*

Example 3.5 *The density function of the gamma distribution appears to be positively skewed. Demonstrate that this is true and illustrate with graphs.*

[1] It would more accurate to call these items the "coefficient of skewness" and "coefficient of kurtosis" because there are other quantities that also measure asymmetry and flatness. The simpler expressions will be used in this note.

From *LMA*, the first three raw moments of the gamma distribution are $\alpha\theta$, $\alpha(\alpha+1)\theta^2$, and $\alpha(\alpha+1)(\alpha+2)\theta^3$. From Equation (3.3) the variance is $\alpha\theta^2$ and from the solution to Exercise 6 the third central moment is $2\alpha\theta^3$. Therefore, the skewness is $2\alpha^{-1/2}$. Because α must be positive, the skewness is always positive. Also, as α decreases, the skewness increases.

Consider the following two gamma distributions. One has parameters $\alpha = 0.5$ and $\theta = 100$ while the other has $\alpha = 5$ and $\theta = 10$. These have the same mean, but their skewness coefficients are 2.83 and 0.89 respectively. The figure below demonstrates the difference.



Densities of f(x)∼gamma(0.5,100) and g(x)∼gamma(5,10)

□

Exercise 7 *Calculate the standard deviation, skewness, and kurtosis for each of the six models. It may help to note that Model 2 is a Pareto distribution and the density function in the continuous part of Model 4 is an exponential distribution. Formulas that may help with calculations for these models appear in LMA.*

Note that when calculating the standard deviation for Model 6 in Exercise 7 the result is the sample standard deviation using n (as opposed to the more commonly used $n-1$) in the denominator. A justification for $n-1$ is provided in the Course 4 reading. Finally, it should be noted that when calculating moments, it is possible that the integral or sum will not exist (as is the case for the third and fourth moments for Model 2). For the models we typically encounter, the integrand and summand are non-negative and so failure to exist implies that the required limit that gives the integral or sum is infinity. See Example 4.15 on Page 31 for an illustration.

Exercise 8 *(*) A random variable has a mean and a coefficient of variation of 2. The third raw moment is 136. Determine the skewness.*

Exercise 9 *(*) Determine the skewness of a gamma distribution that has a coefficient of variation of 1.*

Definition 3.6 *For a given value of d, the* ***excess loss variable*** *is $Y = X - d$ given that $X > d$. Its expected value, $E(Y) = E(X - d|X > d)$ is called the* ***mean excess loss function.*** *It is usually denoted $e(d)$ or $e_X(d)$. Other names for this expectation are* ***mean residual life function*** *and* ***complete expectation of life****. When the latter terminology is used the commonly used symbol is $\mathring{e}_d$.*

This variable could also be called a **left truncated and shifted variable**. It is left truncated because observations below d are discarded. It is shifted because d is subtracted from the remaining values. When X is a payment variable, the mean excess loss is the expected amount paid, given that there has been a payment in excess of a deductible of d. When X is the age at death, the mean excess loss is the expected remaining time until death, given that the person is alive at age d. The kth moment of the excess loss variable is determined from

$$\begin{aligned} e_X^k(d) &= \frac{\int_d^\infty (x-d)^k f(x)dx}{1-F(d)}, \text{ if the random variable is continuous} \\ &= \frac{\sum_{x_i > d}(x_i - d)^k p(x_i)}{1-F(d)}, \text{ if the random variable is discrete.} \end{aligned} \tag{3.4}$$

$e_X^k(d)$ is defined only if the integrand or sum converges. There is a particularly convenient formula for calculating the first moment. The development is given below for the continuous version, but the result holds for all random variables. The second line is based on an integration by parts where the antiderivative of $f(x)$ is taken as $-S(x)$.

$$\begin{aligned} e_X(d) &= \frac{\int_d^\infty (x-d)f(x)dx}{1-F(d)} \\ &= \frac{-(x-d)S(x)|_d^\infty + \int_d^\infty S(x)dx}{S(d)} \\ &= \frac{\int_d^\infty S(x)dx}{S(d)}. \end{aligned} \tag{3.5}$$

Exercise 10 *Determine the mean excess loss function for each of the first four model random variables. Compare the functions for Models 1, 2, and 4.*

Exercise 11 *(*) For two random variables, X and Y, $e_Y(30) = e_X(30) + 4$. X has a uniform distribution on the interval from 0 to 100. Y has a uniform distribution on the interval from 0 to w. Determine w.*

Exercise 12 *(*) The random variable X has density function $f(x) = \lambda^{-1}e^{-x/\lambda}$, $x, \lambda > 0$. Determine $e(\lambda)$, the mean residual life function evaluated at λ.*

Definition 3.7 *The* ***left censored and shifted variable*** *is*

$$Y = (X-d)_+ = \begin{cases} 0, & X < d \\ X - d, & X \geq d. \end{cases}$$

It is left censored because values below d are not ignored, but are set equal to 0. There is no standard name or symbol for the moments of this variable. For dollar events, the distinction between the excess loss variable and the left censored and shifted variable is one of "per payment" versus "per loss." In the former situation, the variable exists only when a payment is made. The latter variable takes on the value 0 whenever a loss produces no payment. The moments can be calculated from

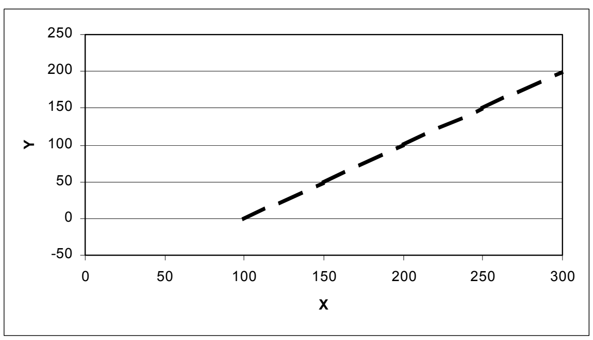
$$\begin{aligned} E[(X-d)_+^k] &= \int_d^\infty (x-d)^k f(x)dx, \text{ if the random variable is continuous} \\ &= \sum_{x_i>d} (x_i-d)^k p(x_i), \text{ if the random variable is discrete.} \end{aligned} \tag{3.6}$$

It should be noted that

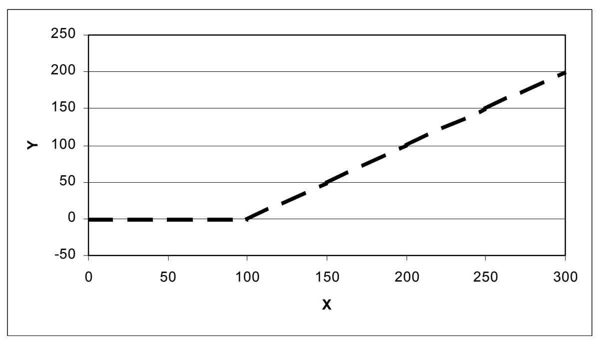
$$E[(X-d)_+^k] = e^k(d)[1-F(d)]. \tag{3.7}$$

Example 3.8 *Construct graphs to illustrate the difference between the excess loss variable and the left censored and shifted variable.*

The two graphs below plot the modified variable Y as a function of the unmodified variable X. The only difference is that for X values below 100, the variable is undefined while for the left censored and shifted variable it is set equal to zero.



Excess Loss Variable



Left Censored and Shifted Variable

□

The next definition provides a complementary function to the excess loss.

Definition 3.9 *The **limited loss variable** is*

$$Y = X \wedge u = \begin{cases} X, & X < u \\ u, & X \geq u. \end{cases}$$

*Its expected value, $E[X \wedge u]$, is called the **limited expected value**.*

This variable could also be called the **right censored variable**. It is right censored because values above u are set equal to u. An insurance phenomenon that relates to this variable is the existence of a policy limit that sets a maximum on the benefit to be paid. Note that $(X-d)_+ + (X \wedge d) = X$. That is, buying one policy with a limit of d and another with a deductible of d is equivalent to buying full coverage. This is illustrated in the following picture.

Limit of 100 plus deductible of 100 equals full coverage

The most direct formulas for the kth moment of the limited loss variable are

$$\begin{aligned} E[(X \wedge u)^k] &= \int_{-\infty}^{u} x^k f(x)dx + u^k[1 - F(u)], \text{ if the random variable is continuous} \\ &= \sum_{x_i \le u} x_i^k p(x_i) + u^k[1 - F(u)], \text{ if the random variable is discrete.} \end{aligned} \tag{3.8}$$

Another interesting formula is derived as follows.

$$\begin{aligned} E[(X \wedge u)^k] &= \int_{-\infty}^{0} x^k f(x)dx + \int_{0}^{u} x^k f(x)dx + u^k[1 - F(u)] \\ &= x^k F(x)\Big|_{-\infty}^{0} - \int_{-\infty}^{0} kx^{k-1}F(x)dx - x^k S(x)\Big|_{0}^{u} + \int_{0}^{u} kx^{k-1}S(x)dx + u^k S(u) \\ &= -\int_{-\infty}^{0} kx^{k-1}F(x)dx + \int_{0}^{u} kx^{k-1}S(x)dx \end{aligned} \tag{3.9}$$

where the second line uses integration by parts. For $k = 1$, we have

$$E(X \wedge u) = -\int_{-\infty}^{0} F(x)dx + \int_{0}^{u} S(x)dx.$$

The corresponding formula for discrete random variables is not particularly interesting. The limited expected value also represents the expected dollar saving per incident when a deductible is imposed. The kth limited moment of many common distributions is presented in *LMA*. The next exercise asks you to develop a relationship between the three first moments introduced previously.

Exercise 13 *Show that the following relationship holds:*

$$E(X) = e(d)S(d) + E(X \wedge d). \tag{3.10}$$

Exercise 14 *Determine the limited expected value function for each of the first four model random variables. Do this using both Equation (3.8) and Equation (3.10). For Models 1 and 2 also obtain the function using Equation (3.9).*

Exercise 15 *(*) Which of the following statements are true?*

1. *The mean residual life function for an empirical distribution is continuous.*
2. *The mean residual life function for an exponential distribution is constant.*
3. *If it exists, the mean residual life function for a Pareto distribution is decreasing.*

Exercise 16 *(*) Losses have a Pareto distribution with* $\alpha = 0.5$ *and* $\theta = 10{,}000$. *Determine the mean residual life at 10,000.*

Exercise 17 *(*) Forty losses have been observed. Sixteen are between 1 and 4/3 and those sixteen losses total 20. Ten losses are between 4/3 and 2 with a total of 15. Ten more are between 2 and 4 with a total of 35. The remaining four losses are greater than 4. Using the empirical model based on these observations, determine* $E(X \wedge 2)$.

Exercise 18 *(*) A sample of size 2000 contains 1700 observations that are no greater than 6000, 30 that are greater than 6000 but no greater than 7000, and 270 that are greater than 7000. The total amount of the 30 observations that are between 6000 and 7000 is 200,000. The value of* $E(X \wedge 6000)$ *for the empirical distribution associated with these observations is 1810. Determine* $E(X \wedge 7000)$ *for the empirical distribution.*

Exercise 19 *Define a right truncated variable and provide a formula for its* k*th moment.*

3.2 Percentiles

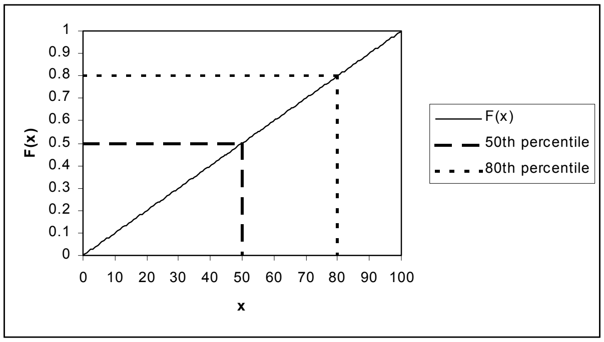
One other value of interest that may be derived from the distribution function is the percentile function. It is the inverse of the distribution function, but because this quantity is not well-defined, an arbitrary definition must be created.

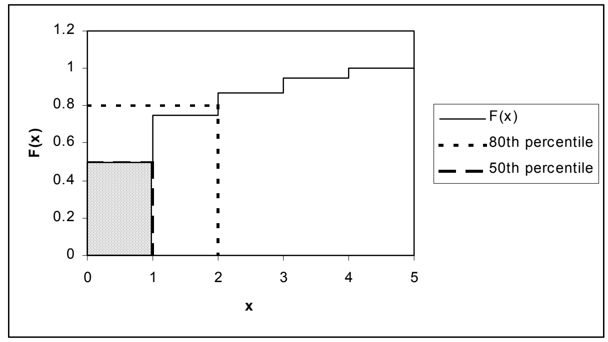
Definition 3.10 *The* ***100pth percentile*** *of a random variable is any value* π_p *such that* $F(\pi_p-) \leq p \leq F(\pi_p)$. *The 50th percentile,* $\pi_{.5}$ *is called the* ***median***.

If the distribution function has a value of p for one and only one x value, then the percentile is uniquely defined. In addition, if the distribution function jumps from a value below p to a value above p, then the percentile is at the location of the jump. The only time the percentile is not uniquely defined is when the distribution function is constant at a value of p over a range of values. In that case, any value in that range can be used as the percentile.

Example 3.11 *Determine the 50th and 80th percentiles for models 1 and 3.*

For Model 1, the pth percentile can be obtained from $p = F(\pi_p) = 0.01\pi_p$ and so $\pi_p = 100p$ and in particular, the requested percentiles are 50 and 80. For Model 3 the distribution function equals 0.5 for all $0 \le x < 1$ and so all such values can be the 50th percentile. For the 80th percentile, note that at $x = 2$ the distribution function jumps from 0.75 to 0.87 and so $\pi_{.8} = 2$. The two graphs below illustrate this process.





Percentiles for Model 1 Percentiles for Model 3

Exercise 20 *Determine the 50th and 80th percentiles for models 2, 4, 5, and 6.*

Exercise 21 *(*) Losses have a Pareto distribution with parameters α and θ. The tenth percentile is $\theta - k$. The ninetieth percentile is $5\theta - 3k$. Determine the value of α.*

Exercise 22 *(*) Losses have a Weibull distribution with parameters τ and θ. The 25th percentile is 1000 and the 75th percentile is 100,000. Determine the value of τ.*

3.3 Moment generating functions and sums of random variables

An insurance company rarely insures only one person. The total claims paid on all policies is the sum of all payments. Thus it is useful to be able to determine properties of $S_k = X_1 + \cdots + X_k$. For this situation we have the following results, given without proof. The first result is a version of the Central Limit Theorem.

Theorem 3.12 *For a random variable S_k as defined above, $E(S_k) = E(X_1) + \cdots + E(X_k)$. Also, if $X_1, \ldots, X_k$ are independent, $Var(S_k) = Var(X_1) + \cdots + Var(X_k)$. If the random variables $X_1, X_2, \ldots$ are independent and their first two moments meet certain conditions, $\lim_{k\to\infty}[S_k - E(S_k)]/\sqrt{Var(S_k)}$ has a normal distribution with mean 0 and variance 1.*

Obtaining the distribution or density function of S_k is usually very difficult. However, there are a few cases where it is simple. To determine those cases we need the moment generating function.

Definition 3.13 *For a random variable X, the moment generating function is $m_X(t) = E(e^{tX})$ for all t for which the expected value exists.*

For us, the value of this function is not that it generates moments but that there is a one-to-one correspondence between a random variable's distribution function and its moment generating function (that is, two random variables with different distribution functions cannot have the same moment generating function). The following result aids in working with sums of random variables.

Theorem 3.14 *Let $S_k = X_1 + \cdots + X_k$ where the random variables in the sum are independent. Then $m_{S_k}(t) = \prod_{j=1}^{k} m_{X_j}(t)$ provided all the component moment generating functions exist.*

Proof: We use the fact that the expected product of independent random variables is the product of the individual expectations. Then,

$$\begin{aligned} m_{S_k}(t) &= E(e^{tS_k}) = E[e^{t(X_1+\cdots+X_k)}] \\ &= \prod_{j=1}^{k} E(e^{tX_j}) = \prod_{j=1}^{k} m_{X_j}(t). \end{aligned}$$

□

Example 3.15 *Show that the sum of independent gamma random variables, each with same value of θ, has a gamma distribution.*

The moment generating function of a gamma variable is

$$\begin{aligned} E(e^{tX}) &= \frac{\int_0^\infty e^{tx} x^{\alpha-1} e^{-x/\theta} dx}{\Gamma(\alpha)\theta^\alpha} \\ &= \frac{\int_0^\infty x^{\alpha-1} e^{-x(-t+1/\theta)} dx}{\Gamma(\alpha)\theta^\alpha} \\ &= \frac{\int_0^\infty y^{\alpha-1} (-t+1/\theta)^{-\alpha} e^{-y} dy}{\Gamma(\alpha)\theta^\alpha} \\ &= \frac{\Gamma(\alpha)(-t+1/\theta)^{-\alpha}}{\Gamma(\alpha)\theta^\alpha} = \left(\frac{1}{1-\theta t}\right)^\alpha. \end{aligned}$$

Now let X_j have a gamma distribution with parameters α_j and θ. Then the moment generating function of the sum is

$$m_{S_k}(t) = \prod_{j=1}^{k} \left(\frac{1}{1-\theta t}\right)^{\alpha_j} = \left(\frac{1}{1-\theta t}\right)^{\alpha_1+\cdots+\alpha_k}$$

which is the moment generating function of a gamma distribution with parameters $\alpha_1 + \cdots + \alpha_k$ and θ. □

Exercise 23 *(*) A portfolio contains 16 independent risks, each with a gamma distribution with parameters $\alpha = 1$ and $\theta = 250$. Give an expression using the incomplete gamma function for the probability that the sum of the losses exceeds 6000. Then approximate this probability using the Central Limit Theorem.*

Chapter 4

Classifying and creating distributions

4.1 Introduction

The set of all possible distribution functions is too large to comprehend. Therefore, when searching for a distribution function to use as a model for some random phenomenon, it can be helpful if the field can be narrowed. One division that has already been discussed is the separation into discrete, continuous, and mixed distributions. In most situations it will be obvious which of the three applies. Beyond this, we need more artificial distinctions. The next section describes a split based on the complexity of the model. The following section then looks at the shape of the distribution to distinguish one from another. After that, a few methods of creating additional distributions are introduced. By the end of this chapter, most of the distributions in *LMA* will have been introduced. While this chapter is more about differences from one distribution to another, these distributions have some common elements that are desirable for actuarial models. Among them are:

- The support is a subset of the non-negative real numbers. Most actuarial phenomena are measurements of time or money and as such are rarely negative, although if the random variable of interest is financial gain, negative outcomes are certainly possible. However, financial gain is often the result of the realization of several non-negative variables, some measuring income and some measuring expenses.

- Some distributions are special cases of others. This allows the modeler to choose from models of varying complexity.

- These models are all continuous. While there are discrete actuarial phenomena, they are the subject of Chapter 3 of *Loss Models* and will not be covered in this note.

- They can have one or more modes and a mode may or may not be at zero.

4.2 Complexity

This split has to do with how much information is needed to specify the model, not the difficulty of using the model. A "complex" model may be simple to obtain and work with, but many items are needed to describe it. Arguments for a simple model include the following.

- With few items required in its specification, it is more likely that each item can be determined more accurately.
- It is more likely to be stable across time and across settings. That is, if the model does well today, it (perhaps with small changes to reflect inflation or similar phenomena) will probably do well tomorrow, and will also do well in other, similar, situations.
- Because data can often be irregular, a simple model may provide necessary smoothing.

Of course, complex models also have advantages.

- With many items required in its specification, it can more closely match reality.
- With many items required in its specification, it can more closely match irregularities in the data.

Another way to express the difference is that simpler models can be estimated more accurately, but the model itself may be too superficial. The principle of parsimony states that the simplest model that adequately reflects reality should be used. The definition of "adequately" will depend on the purpose for which the model is to be used.

In the following subsections, we will move from simpler models to more complex models. There is some difficulty in naming the various classifications because there is not universal agreement on the definitions. With the exception of parametric distributions, the other category names have been created by the author. It should also be understood that these categories do not cover the universe of possible models nor will every model be easy to categorize. These should be considered as qualitative descriptions.

4.2.1 Parametric distributions

These models are simple enough to be specified by a few key numbers.

Definition 4.1 *A **parametric distribution** is a set of distribution functions, each member of which is determined by specifying one or more values called **parameters**. The number of parameters is fixed and finite.*

The most familiar parametric distribution is the normal distribution with parameters μ and σ^2. When values for these two parameters are specified, the distribution function is completely known.

These are the simplest distributions in this subsection, because typically only a small number of values need to be specified. All of the individual distributions in *LMA* are parametric. Within this class, distributions with fewer parameters are simpler than those with more parameters. For much of actuarial modeling work, it is especially convenient if the name of the distribution is unchanged when the random variable is multiplied by a constant. The most common use for this phenomenon is to model the effect of inflation.

Definition 4.2 *A parametric distribution is a **scale distribution** if, when a random variable from that set of distributions is multiplied by a positive constant, the resulting random variable is also in that set of distributions.*

Example 4.3 *Demonstrate that the exponential distribution is a scale distribution.*

According to *LMA*, the distribution function is $F_X(x) = 1 - e^{-x/\theta}$. Let $Y = cX$ where $c > 0$. Then,

$$\begin{aligned} F_Y(y) &= \Pr(Y \le y) \\ &= \Pr(cX \le y) \\ &= \Pr(X \le y/c) \\ &= 1 - e^{-y/c\theta}. \end{aligned}$$

But this is an exponential distribution with parameter $c\theta$. □

Definition 4.4 *A* ***scale parameter*** *is a parameter for a scale distribution that meets two conditions. First, when a member of the scale distribution is multiplied by a positive constant, the scale parameter is multiplied by the same constant. Second, when a member of the scale distribution is multiplied by a positive constant, all other parameters are unchanged.*

Example 4.5 *Demonstrate that the gamma distribution, as defined in* LMA*, has a scale parameter.*

Let X have the gamma distribution and $Y = cX$. Then, using the incomplete gamma notation in *LMA*,

$$\begin{aligned} F_Y(y) &= \Pr(X \le y/c) \\ &= \Gamma(\alpha; y/c\theta) \end{aligned}$$

indicating that Y has a gamma distribution with parameters α and $c\theta$. Therefore, the parameter θ is a scale parameter. □

Many textbooks write the density function for the gamma distribution as

$$f(x) = x^{\alpha-1}e^{-\beta x}\beta^\alpha/\Gamma(\alpha).$$

We have chosen to use the version of the density function that has a scale parameter. When the alternative version is multiplied by c, the parameters become α and β/c. As well, the mean is proportional to θ in our version, while it is proportional to $1/\beta$ in the alternative version. Our version makes it easier to get ballpark estimates of this parameter, although, for the alternative definition, one need only keep in mind that the parameter is inversely proportional to the mean.

It is often possible to recognize a scale parameter from looking at the distribution or density function. In particular, the distribution function would have x always appear as x/θ.

Exercise 24 *Demonstrate that the lognormal distribution as parameterized in* LMA *is a scale distribution, but has no scale parameter. Display an alternative parametrization of this distribution that does have a scale parameter.*

Exercise 25 *Which of Models 1–6 could be considered as members of a parametric distribution? For those that are, name or describe the distribution.*

Exercise 26 *(*) Claims have a Pareto distribution with $\alpha = 2$ and θ unknown. Claims the following year experience 6% uniform inflation. Let r be the ratio of the proportion of claims that will exceed d next year to the proportion of claims that exceed d this year. Determine the limit of r as d goes to infinity.*

4.2.2 Parametric distribution families

A slightly more complex version of a parametric distribution is one in which the number of parameters is finite, but not fixed in advance.

Definition 4.6 *A **parametric distribution family** is a set of parametric distributions that are related in some meaningful way.*

The most common type of parametric distribution family is described in the following example.

Example 4.7 *One type of parametric distribution family is based on a specified parametric distribution. Other members of the family are obtained by setting one or more parameters from the specified distribution equal to a pre-set value or to each other. Demonstrate that the transformed beta family as defined in* LMA *is a parametric distribution family.*

The transformed beta distribution has four parameters. Each of the other named distributions in the family is a transformed beta distribution with certain parameters set equal to one (for example, the Pareto distribution has $\gamma = \tau = 1$) or to each other (the paralogistic distribution has $\tau = 1$ and $\gamma = \alpha$). Note that the number of parameters (ranging from two to four) is not known in advance. There is a subtle difference in definitions. A modeler who uses the transformed beta distribution looks at all four parameters over their range of possible values. A modeler who uses the transformed beta family pays particular attention to the possibility of using special cases such as the Burr distribution. For example, if the former modeler collects some data and decides that $\tau = 1.01$, that will be the value to use. The latter modeler will note that $\tau = 1$ gives a Burr distribution and will likely use that model instead. □

4.2.3 Mixture distributions

By themselves, mixture distributions are no more complex, but later in this subsection we will find a way to increase the complexity level. One motivation for mixing is that the underlying phenomenon may actually be several phenomena that occur with unknown probabilities. For example, a randomly selected dental claim may be from a check-up, from a filling, from a repair (such as a crown), or from a surgical procedure. Because of the differing modes for these possibilities, a mixture model may work well.

Definition 4.8 *A random variable Y is a **k-point mixture**[1] of the random variables $X_1, X_2, \ldots, X_k$ if its cdf is given by*

$$F_Y(y) = a_1 F_{X_1}(y) + a_2 F_{X_2}(y) + \cdots + a_k F_{X_k}(y) \tag{4.1}$$

where all $a_i > 0$ and $a_1 + a_2 + \cdots + a_k = 1$.

[1] In this note a "mixed" distribution refers to one that is partly discrete and continuous, while a "mixture" distribution is one that is a combination of other distributions. The term "mixed" is also commonly used for the second case, but will be avoided here.

This essentially assigns probability a_j to the outcome that Y is a realization of the random variable X_j. Note that if we have 20 choices for a given random variable, a two-point mixture allows us to create over 200 new distributions.[2] This may be sufficient for most modeling situations. Nevertheless, these are still parametric distribution, though perhaps with many parameters.

Example 4.9 *For models involving general liability insurance, the Insurance Services Office has had some success with a mixture of two Pareto distributions. They also found that five parameters were not necessary. The distribution they selected has cdf*

$$F(x) = 1 - a\left(\frac{\theta_1}{\theta_1 + x}\right)^{\alpha} - (1-a)\left(\frac{\theta_2}{\theta_2 + x}\right)^{\alpha+2}.$$

Note that the shape parameters in the two Pareto distributions differ by 2. The second distribution has a lighter tail. This might be a model for frequent, small claims while the first distribution covers large, but infrequent claims. This distribution has only four parameters, bringing some parsimony to the modeling process. □

Exercise 27 *Determine the mean and second moment of the two-point mixture distribution in Example 4.9.*

The solution to this exercise provides general formulas for raw moments of a mixture distribution.

Suppose we do not know how many distributions should be in the mixture. Then the value of k becomes a parameter, as indicated in the following definition.

Definition 4.10 *A* ***variable-component mixture distribution*** *has a distribution function that can be written as*

$$F(x) = \sum_{j=1}^{K} a_j F_j(x),\ \sum_{j=1}^{K} a_j = 1,\ a_j > 0,\ j = 1, \ldots, K,\ K = 1, 2, \ldots$$

These models have been called "semi-parametric" because in complexity they are between parametric models and non-parametric models (see the next subsection). This distinction becomes more important when model selection is discussed in Course 4. When the number of parameters is to be estimated from data, hypothesis tests to determine the appropriate number of parameters become more difficult. When all of the components have the same parametric distribution (but different parameters), the resulting distribution is called a "variable mixture of g" distribution where g stands for the name of the component distribution.

Example 4.11 *Determine the distribution, density, and hazard rate functions for the variable mixture of exponentials distribution.*

A combination of exponential distribution functions can be written

$$F(x) = 1 - a_1 e^{-x/\theta_1} - a_2 e^{-x/\theta_2} - \cdots - a_K e^{-x/\theta_K},\ \sum_{j=1}^{K} a_j = 1,\ a_j, \theta_j > 0,\ j = 1, \ldots, K,\ K = 1, 2, \ldots$$

[2] There are $\binom{20}{2} + 20 = 210$ choices. The extra 20 are included because both distributions could be from the same family, but with different parameter values.

and then the other functions are

$$
\begin{aligned}
f(x) &= a_1\theta_1^{-1}e^{-x/\theta_1} + a_2\theta_2^{-1}e^{-x/\theta_2} + \cdots + a_K\theta_K^{-1}e^{-x/\theta_K} \\
h(x) &= \frac{a_1\theta_1^{-1}e^{-x/\theta_1} + a_2\theta_2^{-1}e^{-x/\theta_2} + \cdots + a_K\theta_K^{-1}e^{-x/\theta_K}}{a_1e^{-x/\theta_1} + a_2e^{-x/\theta_2} + \cdots + a_Ke^{-x/\theta_K}}.
\end{aligned}
$$

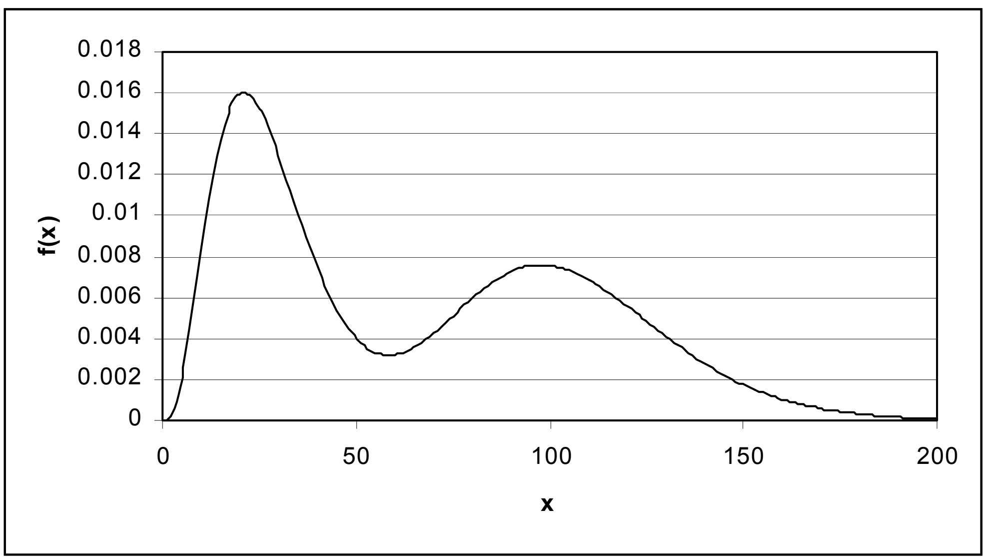
The number of parameters is not fixed nor is it even limited. For example, when $K = 2$ there are three parameters $(a_1, \theta_1, \theta_2)$ noting that a_2 is not a parameter because once a_1 is set the value of a_2 is determined. However, when $K = 4$ there are seven parameters. □

Example 4.12 *Illustrate how a two-point mixture of gamma variables can create a bi-modal distribution.*

Consider a fifty-fifty mixture of two gamma distributions. One has parameters $\alpha = 4$ and $\theta = 7$ (for a mode of 21) and the other has parameters $\alpha = 15$ and $\theta = 7$ (for a mode of 98). The density function is

$$f(x) = 0.5\frac{x^3e^{-x/7}}{3!7^4} + 0.5\frac{x^{14}e^{-x/7}}{14!7^{15}}$$

and a graph appears below.



Two-point mixture of gamma distributions

□

Exercise 28 *Determine expressions for the mean and variance of the mixture of gammas distribution.*

Exercise 29 *Which of Models 1–6 could be considered to be from parametric distribution families? Which could be considered to be from variable-component mixture distributions?*

Exercise 30 *(*) 75% of claims have a normal distribution with a mean of 3000 and a variance of 1,000,000. The remaining 25% have a normal distribution with a mean of 4000 and a variance of 1,000,000. Determine the probability that a randomly selected claim exceeds 5000.*

Exercise 31 *(*) X has a Burr distribution with parameters $\alpha = 1$, $\gamma = 2$, and $\theta = \sqrt{1000}$. Y has a Pareto distribution with parameters $\alpha = 1$ and $\theta = 1000$. Let Z be a mixture of X and Y with equal weight on each component. Determine the median of Z. Let $W = 1.1Z$. Demonstrate that W is also a mixture of a Burr and a Pareto distribution and determine the parameters of W.*

Exercise 32 *(*) Consider three random variables. X is a mixture of a uniform distribution on the interval 0 to 2 and a uniform distribution on the interval 0 to 3. Y is the sum of two random variables, one is uniform on 0 to 2 and the other is uniform on 0 to 3. Z is a normal distribution that has been right censored at 1. Match these random variables with the following descriptions.*

1. *Both the distribution and density functions are continuous.*
2. *The distribution function is continuous but the density function is discontinuous.*
3. *The distribution function is discontinuous.*

4.2.4 Data-dependent distributions

Models 1–5 and many of the examples rely on an associated phenomenon (the random variable) but not on observations of that phenomenon. For example, without having observed any dental claims, I could postulate a lognormal distribution with parameters $\mu = 5$ and $\sigma = 1$. My model may be a poor description of dental claims, but that is a different matter. On the other hand, it is possible to construct models that require data. These models also have parameters, but are often called non-parametric. As will be made clear in Exam 4, model selection issues become even more challenging when the model itself (as opposed to its parameters) are data-dependent.

Definition 4.13 *A* ***data-dependent distribution*** *is at least as complex as the data or knowledge that produced it and the number of "parameters" increases as the number of data points or amount of knowledge increases.*

Essentially, these models have as many (or more) "parameters" than observations in the data set. The empirical distribution as illustrated by Model 6 on Page 14 is a data-dependent distribution. Each data point contributes probability $1/n$ to the probability function, so the n parameters are the n observations in the data set that produced the empirical distribution.

Another example of a data-dependent model is the kernel smoothing model introduced in Course 4. Rather than place a spike of probability $1/n$ at each data point, a continuous density function with area $1/n$ replaces the data point. This piece is centered at the data point so that this model follows the data, but not perfectly. It provides some smoothing versus the empirical distribution. A simple example is given below. It is not expected that Exam 3 candidates should be able to derive this model from data, but like all models in this note, once the distribution, density, or hazard rate functions are known, all model-based calculations can be done.

Example 4.14 *Construct a kernel smoothing model from Model 6 using the uniform kernel and a bandwidth of 2.*

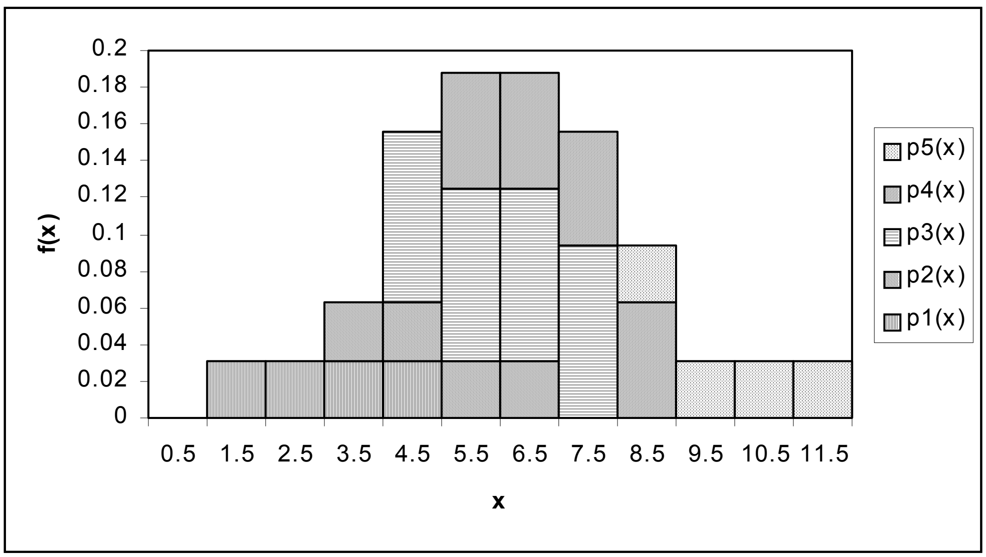
The probability density function is

$$f(x) = \sum_{j=1}^{5} p_6(x_j)K_j(x)$$
$$K_j(x) = \begin{cases} 0, & |x - x_j| > 2 \\ 0.25, & |x - x_j| \leq 2 \end{cases}$$

where the sum is taken over the five points where the original model has positive probability. For example, the first term of the sum is the function

$$p_6(x_1)K_1(x) = \left\{ \begin{array}{ll} 0, & x < 1 \\ 0.03125, & 1 \leq x \leq 5 \\ 0, & x > 5 \end{array} \right\}.$$

The complete density function is the sum of five such functions, which are illustrated in the following graph.



Kernel density distribution

□

Note that both the kernel smoothing model and the empirical distribution can also be written as mixture distributions. The reason these models are classified separately is that the number of components relates to the sample size rather than to the phenomenon and its random variable.

Exercise 33 *Demonstrate that the model in Example 4.14 is a mixture of uniform distributions.*

4.3 Tail weight

The "tail" of a distribution (more properly, the right tail) is that part that reveals probabilities about large values. It is of interest to actuaries because it is the occurrence (or lack thereof) of large values that is most influential on profits. Risky types of insurance such as medical malpractice feature more large claims (relative to the mean) than less risky insurances such as automobile physical damage. Random variables that tend to assign higher probabilities to large values are said to be heavy tailed. Tail weight can be a relative concept (model A has a heavier tail than model B) or an absolute concept (distributions with a certain property are classified as heavy tailed). When choosing models, tail weight can help narrow the choices, or can confirm someone else's choice. For example, when someone models medical malpractice payments with a Pareto distribution, it seems reasonable as the Pareto distribution is regarded as having a heavy tail. Conversely, the light-tailed lognormal distribution may be a reasonable model for dental insurance payments. However, it should be noted that various measures of tail weight need not agree.

4.3.1 Existence of moments

Recall that in the continuous case, the kth raw moment for a random variable that takes on only positive values (like most insurance payment variables) is given by $\int_0^\infty x^k f(x)dx$. Depending on the density function and the value of k, this integral may not exist. If the density function is too large for large values of x, then when multiplied by the large number x^k, the values will be too large for the integral to converge. Thus, existence of all positive moments indicates a light right tail, while existence only for positive moments up to a certain value (or existence of no positive moments at all) indicates a heavy right tail.[3]

Example 4.15 *Demonstrate that for the gamma distribution all positive raw moments exist, but for the Pareto distribution they do not.*

For the gamma distribution

$$\begin{aligned}
\mu'_k &= \int_0^\infty x^k \frac{x^{\alpha-1}e^{-x/\theta}}{\Gamma(\alpha)\theta^\alpha}dx \\
&= \int_0^\infty (y\theta)^k \frac{(y\theta)^{\alpha-1}e^{-y}}{\Gamma(\alpha)\theta^\alpha}\theta dy, \text{ making the substitution } y = x/\theta \\
&= \frac{\theta^k}{\Gamma(\alpha)}\Gamma(\alpha+k) < \infty \text{ for all } k > 0.
\end{aligned}$$

[3] In the same manner, existence of all negative moments indicates a light left tail. A feel for the left tail may aid in choosing an appropriate distributional model. The following statements are not always true, but do hold if the density function is monotonic and differentiable near zero.

In particular, $f(0)$ and the slope of the density function near zero are related to the existence of negative moments. Suppose that negative moments exist only for $k > -r$. If $r < 1$, $f(x)$ goes to infinity as $x \to 0$. If $r = 1$, $f(0)$ is a non-negative number. If $1 < r < 2$, $f(0) = 0$ and the slope goes to infinity as $x \to 0$. If $r = 2$, $f(0) = 0$ and the slope at 0 is a non-negative number. If $r > 2$, $f(0) = 0$ and the slope at 0 is 0, so only a small portion of the distribution is near zero.

As an example, the Weibull distribution with $\tau = 0.2$ has been used for workers compensation insurance. The value of r is 0.2 which means that a lot of probability is near 0. Setting θ to produce a mean of 30,000 gives a 28% chance of a claim being less than 1 and a 1% chance of it exceeding 500,000. This may be a reasonable model for large losses (see Section 4.4.6 for a way to use part of a model). In contrast, the lognormal distribution that has the same mean and variance has less than 0.1% probability of being less than 1 and also about 1% probability of exceeding 500,000. The lognormal distribution has all negative moments and thus has $f(x)$ go to zero as x goes to zero.

while for the Pareto distribution

$$\begin{aligned}\mu_k' &= \int_0^\infty x^k \frac{\alpha\theta^\alpha}{(x+\theta)^{\alpha+1}}dx \\ &= \int_\theta^\infty (y-\theta)^k \frac{\alpha\theta^\alpha}{y^{\alpha+1}}dy, \text{ making the substitution } y = x+\theta \\ &= \alpha\theta^\alpha \int_\theta^\infty \sum_{j=0}^k \binom{k}{j} y^{j-\alpha-1}(-\theta)^{k-j}dy \text{ for integral values of } k.\end{aligned}$$

The integral exists only if all of the exponents on y in the sum are less than -1. That is, $j-\alpha-1 < -1$ for all j, or equivalently, $k < \alpha$. Therefore, only some moments exist. □

By this reckoning, the Pareto distribution is said to have a heavier tail than the gamma distribution. A look at the moment formulas in *LMA* reveals which distributions have heavy tails and which do not as indicated by the existence of moments.

4.3.2 Limiting density ratios

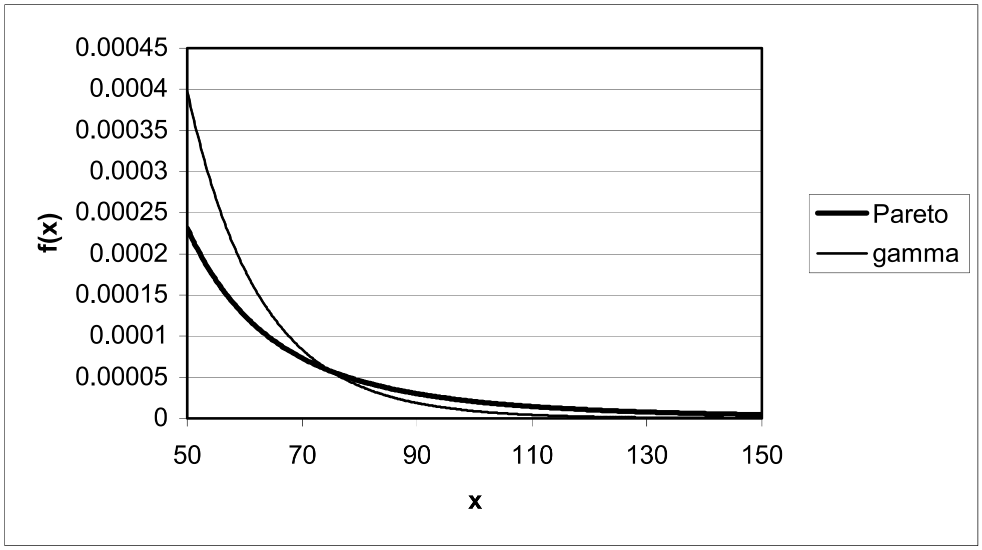
An indication that one distribution has a heavier tail than another is that the ratio of the two density functions should diverge to infinity (with the heavier tailed distribution in the numerator). The same result will obtain if the ratio of the survival functions is used.

Example 4.16 *Demonstrate that the Pareto distribution has a heavier tail than the gamma distribution using the limit of the ratio of their density functions.*

To avoid confusion, the letters τ and λ will be used for the parameters of the gamma distribution instead of the customary α and θ. Then the required limit is

$$\begin{aligned}\lim_{x\to\infty}\frac{f_{\text{Pareto}}(x)}{f_{\text{gamma}}(x)} &= \lim_{x\to\infty}\frac{\alpha\theta^\alpha(x+\theta)^{-\alpha-1}}{x^{\tau-1}e^{-x/\lambda}\lambda^{-\tau}\Gamma(\tau)^{-1}} \\ &= c\lim_{x\to\infty}\frac{e^{x/\lambda}}{(x+\theta)^{\alpha+1}x^{\tau-1}} \\ &> c\lim_{x\to\infty}\frac{e^{x/\lambda}}{(x+\theta)^{\alpha+\tau}}\end{aligned}$$

and, either by application of L'Hôpital's rule or by remembering that exponentials go to infinity faster than polynomials, the limit is infinity. The following graph shows a portion of the density functions for a Pareto distribution with parameters $\alpha = 3$ and $\theta = 10$ and a gamma distribution with parameters $\alpha = 1/3$ and $\theta = 15$. Both distributions have a mean of 5 and a variance of 75. The graph is consistent with the algebraic derivation.



Tails of gamma and Pareto distributions

□

4.3.3 Hazard rate and mean residual life patterns

The nature of the hazard rate function also reveals information about the tail of the distribution. If the hazard rate function is decreasing, then at large values the chance of that value becomes small and the chance of larger values becomes greater. Thus the distribution will have a heavier tail. Conversely, if the hazard rate function is increasing, a lighter tail is expected.

Example 4.17 *Compare the tails of the Pareto and gamma distributions by looking at their hazard rate functions.*

The hazard rate function for the Pareto distribution is

$$h(x) = \frac{f(x)}{S(x)} = \frac{\alpha\theta^\alpha(x+\theta)^{-\alpha-1}}{\theta^\alpha(x+\theta)^{-\alpha}} = \frac{\alpha}{x+\theta}$$

which is decreasing. For the gamma distribution we need to be a bit more clever because there is no closed form expression for $S(x)$. Observe that

$$\frac{1}{h(x)} = \frac{\int_x^\infty f(t)dt}{f(x)} = \frac{\int_0^\infty f(x+y)dy}{f(x)}$$

and so if $f(x+y)/f(x)$ is an increasing function of x for any fixed y, then $1/h(x)$ will be increasing in x and so the random variable will have a decreasing hazard rate. Now, for the gamma distribution

$$\frac{f(x+y)}{f(x)} = \frac{(x+y)^{\alpha-1}e^{-(x+y)/\theta}}{x^{\alpha-1}e^{-x/\theta}} = \left(1+\frac{y}{x}\right)^{\alpha-1} e^{-y/\theta}$$

which is strictly increasing in x provided $\alpha < 1$ and strictly decreasing in x if $\alpha > 1$. By this measure, some gamma distributions have a heavy tail (those with $\alpha < 1$) and some have a light tail. Note that when $\alpha = 1$ we have the exponential distribution and a constant hazard rate. □

The mean residual life also gives information about tail weight. If the mean residual life function is increasing in d, then at large values, the expected outcome is much larger and thus probability is moved to the right, indicating a heavier tail than a model where the mean residual life function is decreasing or is increasing at a slower rate. It is proved in *Loss Models* (pages 89–91) that if a random variable has a decreasing hazard rate, then it must have an increasing mean residual life function and similarly, an increasing hazard rate implies a decreasing mean residual life function.

Example 4.18 *Compare the tails of the Pareto and gamma distributions by looking at their mean residual life functions.*

Based on the previous example, the answer is already known. From the definition, for the Pareto distribution,

$$e(x) = \frac{\int_x^\infty (y-x)\alpha\theta^\alpha(y+\theta)^{-\alpha-1}dy}{\theta^\alpha(x+\theta)^{-\alpha}} = \frac{\alpha\theta^\alpha \frac{(x+\theta)^{-\alpha+1}}{\alpha(\alpha-1)}}{\theta^\alpha(x+\theta)^{-\alpha}} = \frac{x+\theta}{\alpha-1},\ \alpha > 1$$

which is increasing. For the gamma distribution the function is more complicated and it is best to rely on the relationship to the hazard rate function. In *Loss Models* (page 91) it is shown that for the gamma distribution $\lim_{x\to\infty} e(x) = \theta$. For the heavy tailed case ($\alpha < 1$) we know the mean residual life is strictly increasing, but unlike the Pareto distribution, it increases to a finite number, rather than to infinity. □

Exercise 34 *Using the methods in this section (except for the mean residual life), compare the tail weight of the Weibull and Inverse Weibull distributions.*

4.4 Creating new distributions

4.4.1 Introduction

This section indicates how new parametric distribution can be created from existing ones. Many of the distributions in *LMA* were created this way.

4.4.2 Multiplication by a constant

This transformation is equivalent to applying inflation uniformly across all loss levels and is known as a change of scale. For example, if this year's losses are given by the random variable X, then uniform inflation of 5% indicates that next year's losses can be modeled with the random variable $Y = 1.05X$.

Theorem 4.19 *Let X be a continuous random variable with pdf $f_X(x)$ and cdf $F_X(x)$. Let $Y = \theta X$ with $\theta > 0$. Then*

$$F_Y(y) = F_X(y/\theta), \quad f_Y(y) = \frac{1}{\theta} f_X(y/\theta).$$

Proof:

$$F_Y(y) = \Pr(Y \le y) = \Pr(\theta X \le y) = \Pr(X \le y/\theta) = F_X(y/\theta)$$

$$f_Y(y) = \frac{d}{dy}F_Y(y) = \frac{1}{\theta}f_X(y/\theta).$$

□

Corollary 4.20 *θ is a scale parameter for the random variable Y.* □

The following example illustrates this process.

Example 4.21 *Let X have pdf $f(x) = e^{-x}$, $x > 0$. Determine the cdf and pdf of $Y = \theta X$.*

$$\begin{aligned} F_X(x) &= 1 - e^{-x}, \quad F_Y(y) = 1 - e^{-y/\theta} \\ f_Y(y) &= \frac{1}{\theta}e^{-y/\theta}. \end{aligned}$$

We recognize this as the exponential distribution. □

Exercise 35 *Let X have cdf $F_X(x) = 1 - (1+x)^{-\alpha}$, $x, \alpha > 0$. Determine the pdf and cdf of $Y = \theta X$.*

Exercise 36 *(*) 100 observed claims in 1995 were arranged as follows. 42 were between 0 and 300, 3 were between 300 and 350, 5 were between 350 and 400, 5 were between 400 and 450, 0 were between 450 and 500, 5 were between 500 and 600, and the remaining 40 were above 600. For the next three years, all claims are inflated by 10% per year. Based on the empirical distribution from 1995, determine a range for the probability that a claim exceeds 500 in 1998 (there is not enough information to determine the probability exactly).*

4.4.3 Raising to a power

Theorem 4.22 *Let X be a continuous random variable with pdf $f_X(x)$ and cdf $F_X(x)$ with $F_X(0) = 0$. Let $Y = X^{1/\tau}$. Then if $\tau > 0$,*

$$F_Y(y) = F_X(y^\tau), \quad f_Y(y) = \tau y^{\tau-1} f_X(y^\tau), \qquad y > 0$$

while if $\tau < 0$,

$$F_Y(y) = 1 - F_X(y^\tau), \quad f_Y(y) = -\tau y^{\tau-1} f_X(y^\tau). \tag{4.2}$$

Proof: If $\tau > 0$

$$F_Y(y) = \Pr(X \le y^\tau) = F_X(y^\tau)$$

while if $\tau < 0$

$$F_Y(y) = \Pr(X \ge y^\tau) = 1 - F_X(y^\tau).$$

The pdf follows by differentiation. □

It is more common to keep parameters positive and so when τ is negative, create a new parameter $\tau^* = -\tau$. Then (4.2) becomes

$$F_Y(y) = 1 - F_X(y^{-\tau^*}), \quad f_Y(y) = \tau^* y^{-\tau^*-1} f_X(y^{-\tau^*}).$$

Drop the * for future use of this positive parameter.

Definition 4.23 *When raising a distribution to a power, if $\tau > 0$ the resulting distribution is called* ***transformed****, if $\tau = -1$ it is called* ***inverse****, and if $\tau < 0$ (but is not -1) it is called* ***inverse transformed****. To create the distributions in* LMA *and to retain θ as a scale parameter, the base distribution should be raised to a power before being multiplied by θ.*

Example 4.24 *Suppose X has the exponential distribution. Determine the cdf of the inverse, transformed, and inverse transformed exponential distributions.*

The inverse exponential distribution with no scale parameter has cdf

$$F(y) = 1 - [1 - e^{-1/y}] = e^{-1/y}.$$

With the scale parameter added it is $F(y) = e^{-\theta/y}$.

The transformed exponential distribution with no scale parameter has cdf

$$F(y) = 1 - \exp(-y^{\tau}).$$

With the scale parameter added it is $F(y) = 1 - \exp[-(y/\theta)^{\tau}]$. This distribution is more commonly known as the **Weibull distribution**.

The inverse transformed exponential distribution with no scale parameter has cdf

$$F(y) = 1 - [1 - \exp(-y^{-\tau})] = \exp(-y^{-\tau}).$$

With the scale parameter added it is $F(y) = \exp[-(\theta/y)^{\tau}]$. This distribution is the **inverse Weibull**. □

Exercise 37 *Let X have the Pareto distribution. Determine the cdf of the transformed, inverse, and inverse transformed distributions. Check* LMA *to determine if any of these distributions have special names.*

Exercise 38 *Let X have the loglogistic distribution. Demonstrate that the inverse distribution also has the loglogistic distribution. Therefore there is no need to identify a separate inverse loglogistic distribution.*

Another base distribution has pdf $f(x) = x^{\alpha-1}e^{-x}/\Gamma(\alpha)$. When a scale parameter is added, this becomes the **gamma distribution.** It has inverse and transformed versions that can be created using the results in this section. Unlike the distributions introduced to this point, this one does not have a closed form cdf. The best we can do is define notation for the function.

Definition 4.25 *The* ***incomplete gamma function*** *with parameter $\alpha > 0$ is denoted and defined by*

$$\Gamma(\alpha; x) = \frac{1}{\Gamma(\alpha)} \int_0^x t^{\alpha-1}e^{-t}\,dt$$

while the ***gamma function*** *is denoted and defined by*

$$\Gamma(\alpha) = \int_0^{\infty} t^{\alpha-1}e^{-t}\,dt.$$

In addition, $\Gamma(\alpha) = (\alpha - 1)\Gamma(\alpha - 1)$ and for positive integer values of n, $\Gamma(n) = (n-1)!$. *LMA* provides details on numerical methods of evaluating these quantities. Furthermore, these functions are built into most spreadsheet programs and many statistical and numerical analysis programs.

4.4.4 Exponentiation

Theorem 4.26 *Let X be a continuous random variable with pdf $f_X(x)$ and cdf $F_X(x)$ with $f_X(x) > 0$ for all real x. Let $Y = \exp(X)$. Then, for $y > 0$,*

$$F_Y(y) = F_X(\ln y), \quad f_Y(y) = \frac{1}{y} f_X(\ln y).$$

Proof: $F_Y(y) = \Pr(e^X \leq y) = \Pr(X \leq \ln y) = F_X(\ln y)$. □

Example 4.27 *Let X have the normal distribution with mean μ and variance σ^2. Determine the cdf and pdf of $Y = e^X$.*

$$F_Y(y) = \Phi\left(\frac{\ln y - \mu}{\sigma}\right)$$

$$f_Y(y) = \frac{1}{y\sigma}\phi\left(\frac{\ln y - \mu}{\sigma}\right) = \frac{1}{y\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{\ln y - \mu}{\sigma}\right)^2\right].$$

□

We could try to add a scale parameter by creating $W = \theta Y$, but this adds no value, as is demonstrated in Exercise 39. This example created the **lognormal** distribution (the name has stuck even though expnormal would seem more descriptive).

Exercise 39 *Let Y have the lognormal distribution with parameters μ and σ. Let $Z = \theta Y$. Show that Z also has the lognormal distribution and therefore the addition of a third parameter has not created a new distribution.*

Exercise 40 *(*) Let X have a Pareto distribution with parameters α and θ. Let $Y = \ln(1 + X/\theta)$. Determine the name of the distribution of Y and its parameters.*

4.4.5 Mixing

The concept of mixing can be extended from mixing a finite number of random variables to mixing an uncountable number. In the following Theorem, the pdf $f_\Theta(\theta)$ plays the role of the coefficients a_j in the k-point mixture.

Theorem 4.28 *Let X have pdf $f_{X|\Theta}(x|\theta)$ and cdf $F_{X|\Theta}(x|\theta)$ where θ is a parameter of X. Let θ be a realization of the random variable Θ with pdf $f_\Theta(\theta)$. Then the unconditional pdf of X is*

$$f_X(x) = \int f_{X|\Theta}(x|\theta) f_\Theta(\theta)\, d\theta$$

where the integral is taken over all values of θ with positive probability. The resulting distribution is a ***mixture distribution****. The distribution function can be determined from*

$$\begin{aligned} F_X(x) &= \int_{-\infty}^{x} \int f_{X|\Theta}(y|\theta) f_\Theta(\theta) d\theta dy \\ &= \int \int_{-\infty}^{x} f_{X|\Theta}(y|\theta) f_\Theta(\theta) dy d\theta \\ &= \int F_{X|\Theta}(x|\theta) f_\Theta(\theta) d\theta. \end{aligned}$$

Moments of the mixture distribution can be found from

$$E(X^k) = E[E(X^k|\Theta)]$$

and, in particular

$$Var(X) = E[Var(X|\Theta)] + Var[E(X|\Theta)].$$

Proof: The integrand is, by definition, the joint density of X and Θ. The integral is then the marginal density. For the expected value (assuming the order of integration can be reversed),

$$\begin{aligned} E(X^k) &= \int \int x^k f_{X|\Theta}(x|\theta) f_\Theta(\theta)\, d\theta dx \\ &= \int \left[\int x^k f_{X|\Theta}(x|\theta) dx \right] f_\Theta(\theta) d\theta \\ &= \int E(X^k|\theta) f_\Theta(\theta) d\theta \\ &= E[E(X^k|\Theta)]. \end{aligned}$$

For the variance,

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

□

The following example shows how a familiar distribution may be obtained by mixing.

Example 4.29 *Let $X|\Theta$ have an exponential distribution with parameter $1/\Theta$. Let Θ have a gamma distribution with parameters α and β (using β in place of θ). Determine the unconditional distribution of X.*

We have,

$$\begin{aligned} f_X(x) &= \frac{1}{\Gamma(\alpha)\beta^\alpha}\int_0^\infty \theta e^{-\theta x}\theta^{\alpha-1}e^{-\theta/\beta}d\theta \\ &= \frac{1}{\Gamma(\alpha)\beta^\alpha}\int_0^\infty \theta^\alpha e^{-\theta(x+1/\beta)}d\theta \\ &= \frac{1}{\Gamma(\alpha)\beta^\alpha}\frac{\Gamma(\alpha+1)}{(x+1/\beta)^{\alpha+1}} \\ &= \frac{\alpha\beta}{(1+x\beta)^{\alpha+1}}. \end{aligned}$$

This is a Pareto distribution with the usual (as defined in *LMA*) parameter θ replaced with $1/\beta$. It may appear that a heavy tailed distribution has been created by mixing two light tailed distributions. However, the exponential mean is $1/\Theta$ and so the mean has an inverse gamma distribution, which is heavy tailed. In fact its tail is not unlike that of Pareto distribution, making the result less surprising. □

The following example is taken from Hayne ("Extended Service Contracts," *Proceedings of the Casualty Actuarial Society*, **LXXXI** (1994), pp. 243–302). It illustrates how this type of mixture distribution can arise. In particular, continuous mixtures are often used to provide a model for parameter uncertainty. That is, the exact value of a parameter is not known, but a probability density function can be elucidated to describe possible values of that parameter.

Example 4.30 *In the valuation of warranties on automobiles it is important to recognize that the number of miles driven varies from driver to driver. It is also the case that for a particular driver, the number of miles varies from year to year. Suppose the number of miles for a randomly selected driver has the inverse Weibull distribution but that the year to year variation in the scale parameter has the transformed gamma distribution with the same value for τ. What is the distribution for the number of miles driven in a randomly selected year by a randomly selected driver?*

Using the parametrizations from *LMA*, the inverse Weibull for miles driven in a year has parameters Θ and τ while the transformed gamma distribution for the scale parameter Θ has parameters

τ, γ and α (γ has been substituted for θ). The marginal density is

$$\begin{aligned}
f(x) &= \int_0^\infty \frac{\tau\theta^\tau}{x^{\tau+1}} e^{-(\theta/x)^\tau} \frac{\tau\theta^{\tau\alpha-1}}{\gamma^{\tau\alpha}\Gamma(\alpha)} e^{-(\theta/\gamma)^\tau}\, d\theta \\
&= \frac{\tau^2}{\gamma^{\tau\alpha}\Gamma(\alpha)x^{\tau+1}} \int_0^\infty \theta^{\tau+\tau\alpha-1} \exp[-\theta^\tau(x^{-\tau}+\gamma^{-\tau})]\, d\theta \\
&= \frac{\tau^2}{\gamma^{\tau\alpha}\Gamma(\alpha)x^{\tau+1}} \int_0^\infty \{y^{1/\tau}(x^{-\tau}+\gamma^{-\tau})^{-1/\tau}\}^{\tau+\tau\alpha-1} e^{-y} \\
&\quad \times y^{\tau^{-1}-1}\tau^{-1}(x^{-\tau}+\gamma^{-\tau})^{-1/\tau}\, dy \\
&= \frac{\tau}{\gamma^{\tau\alpha}\Gamma(\alpha)x^{\tau+1}(x^{-\tau}+\gamma^{-\tau})^{\alpha+1}} \int_0^\infty y^\alpha e^{-y}\, dy \\
&= \frac{\tau\Gamma(\alpha+1)}{\gamma^{\tau\alpha}\Gamma(\alpha)x^{\tau+1}(x^{-\tau}+\gamma^{-\tau})^{\alpha+1}} \\
&= \frac{\tau\alpha\gamma^\tau x^{\tau\alpha-1}}{(x^\tau+\gamma^\tau)^{\alpha+1}}.
\end{aligned}$$

The third line is obtained by the transformation $y = \theta^\tau(x^{-\tau}+\gamma^{-\tau})$. The final line uses the fact that $\Gamma(\alpha+1) = \alpha\Gamma(\alpha)$. Note that this distribution applies to a particular driver. Another driver may have a different Weibull shape parameter τ and as well that driver's Weibull scale parameter Θ may have a different distribution, and in particular, a different mean. □

Exercise 41 *In "Transformed Beta and Gamma Distributions and Aggregate Losses" (Proceedings of the Casualty Actuarial Society,* ***LXX*** *(1983), pp. 156–193), Venter noted that if X has the transformed gamma distribution and its scale parameter θ has an inverse transformed gamma distribution (where the parameter τ is the same in both distributions) the resulting mixture has the transformed beta distribution. Demonstrate that this is true.*

Exercise 42 *(*) Let N have a Poisson distribution with mean Λ. Let Λ have a gamma distribution with mean 1 and variance 2. Determine the unconditional probability that $N = 1$.*

Exercise 43 *(*) Given a value of $\Theta = \theta$, the random variable X has an exponential distribution with hazard rate function $h(x) = \theta$, a constant. The random variable Θ has a uniform distribution on the interval $(1, 11)$. Determine $S_X(0.5)$ for the unconditional distribution.*

Exercise 44 *(*) Let N have a Poisson distribution with mean Λ. Let Λ have a uniform distribution on the interval $(0, 5)$. Determine the unconditional probability that $N \geq 2$.*

4.4.6 Splicing

A final method for creating a new distribution is by splicing. This approach is similar to mixing in that it might be believed that two or more separate processes are responsible for generating the losses. With mixing, the various processes operate on subsets of the population. Once the subset is identified, a simple loss model suffices. For splicing, the processes differ with regard to the loss amount. That is, one model governs the behavior of losses in some interval of possible losses while other models cover the other intervals. The following definition makes this precise.

Definition 4.31 *A **k-component spliced model** has a density function that can be expressed as follows.*

$$f_X(x) = \begin{cases} a_1 f_1(x), & c_0 < x < c_1 \\ a_2 f_2(x), & c_1 < x < c_2 \\ \vdots & \vdots \\ a_k f_k(x), & c_{k-1} < x < c_k. \end{cases}$$

For $j = 1, \ldots, k$, each $a_j > 0$, and each $f_j(x)$ must be a legitimate density function with all probability on the interval (c_{j-1}, c_j). Also, $a_1 + \cdots + a_k = 1$.

Example 4.32 *Demonstrate that Model 5 on Page 11 is a 2-component spliced model.*

The density function is

$$f(x) = \begin{cases} 0.01, & 0 \le x < 50 \\ 0.02, & 50 \le x < 75 \end{cases}$$

and the spliced model is created by letting $f_1(x) = 0.02$, $0 \le x < 50$ which is a uniform distribution on the interval from 0 to 50 and $f_2(x) = 0.04$, $50 \le x < 75$ which is a uniform distribution on the interval from 50 to 75. The coefficients are then $a_1 = 0.5$ and $a_2 = 0.5$. □

It was not necessary to use density functions and coefficients, but this is one way to ensure that the result is a legitimate density function. When using parametric models the motivation for splicing is that the tail behavior may be inconsistent with the behavior for small losses. For example, experience (based on knowledge beyond that available in the current, perhaps small, data set) may indicate that the tail follows the Pareto distribution, but there is a positive mode more in keeping with the lognormal or inverse Gaussian distributions. A second instance is when there is a large amount of data below some value, but a limited amount of information elsewhere. We may want to use the empirical distribution (or a smoothed version of it) up to a certain point and a parametric model beyond that value. The definition given above is appropriate when the break points $c_0, \ldots, c_k$ are known in advance.

Another way to construct a spliced model is to use standard distributions over the range from c_0 to c_k. Let $g_j(x)$ be the jth such density function. Then, in Definition 4.31 replace $f_j(x)$ with $g_j(x)/[G(c_j) - G(c_{j-1})]$. This formulation makes it easier to have the break points become parameters that can be estimated.

Neither approach to splicing ensures that the resulting density function will be continuous (that is, the components will meet at the breakpoints). Such a restriction could be added to the specification.

Example 4.33 *Create a two-component spliced model using an exponential distribution from 0 to c and a Pareto distribution (using γ in place of θ) from c to ∞.*

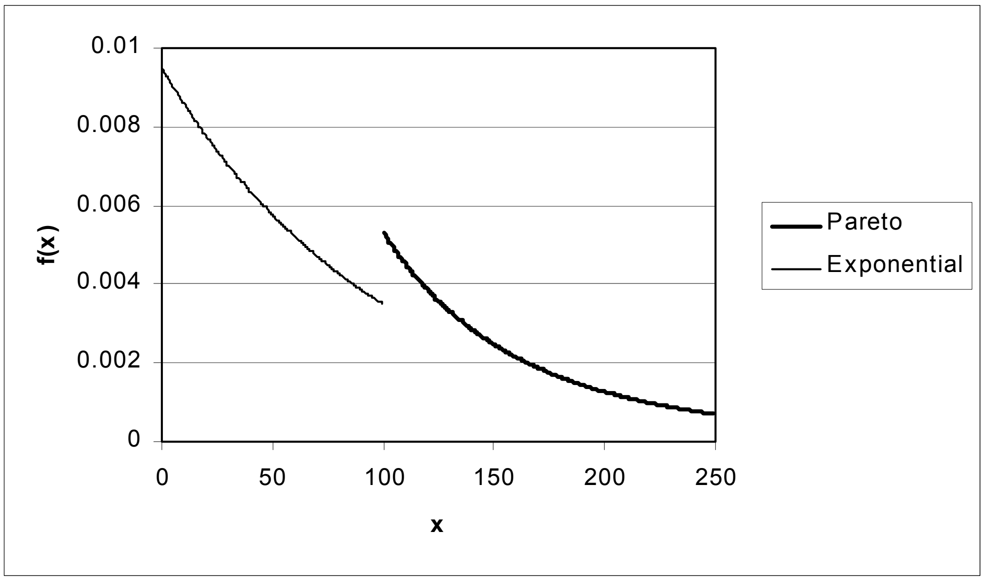
The basic format is

$$f_X(x) = \begin{cases} a_1 \frac{\theta^{-1} e^{-x/\theta}}{1 - e^{-c/\theta}}, & 0 < x < c \\ a_2 \frac{\alpha \gamma^{\alpha} (x+\gamma)^{-\alpha-1}}{\gamma^{\alpha} (c+\gamma)^{-\alpha}}, & c < x < \infty \end{cases}.$$

However, we must force the density function to integrate to 1. All that is needed is to let $a_1 = v$ and $a_2 = 1 - v$. The spliced density function becomes

$$f_X(x) = \begin{cases} v\frac{\theta^{-1}e^{-x/\theta}}{1-e^{-c/\theta}}, & 0 < x < c \\ (1-v)\frac{\alpha(c+\gamma)^\alpha}{(x+\gamma)^{\alpha+1}}, & c < x < \infty \end{cases}, \ \theta, \alpha, \gamma, c > 0,\ 0 < v < 1.$$

The following graph illustrates this density function using the values $c = 100$, $v = 0.6$, $\theta = 100$, $\gamma = 200$, and $\alpha = 4$. It is clear that this density is not continuous.



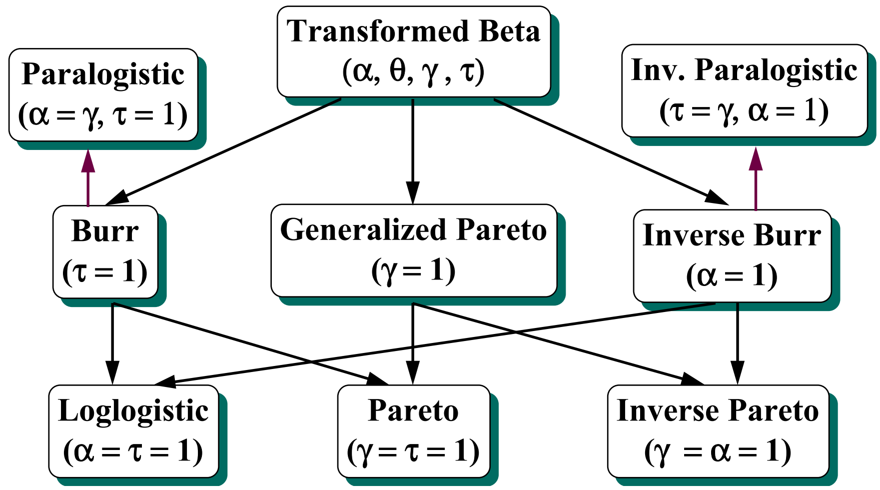
Two-component spliced density

□

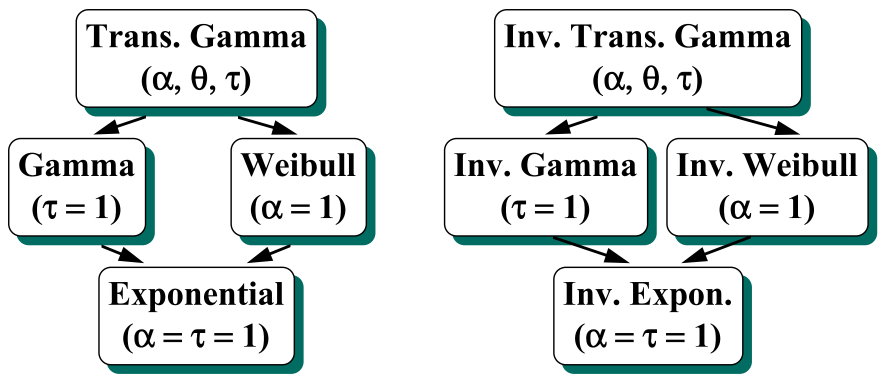
Exercise 45 *Write the density function for a two-component spliced model in which the density function is proportional to a uniform density over the interval from* 0 *to* 1000 *and is proportional to an exponential density function from* 1000 *to* ∞. *Ensure that the resulting density function is continuous.*

4.4.7 Two parametric families

As noted when defining parametric families, many of the distributions presented in this section and in *LMA* are special cases of others. For example, a Weibull distribution with $\tau = 1$ and θ arbitrary is an exponential distribution. Through this process, many of our distributions can be organized into groupings as illustrated in the following two figures. The transformed beta family indicates two special cases of a different nature. The paralogistic and inverse paralogistic distributions are created by setting the two non-scale parameters of the Burr and inverse Burr distributions equal to each other rather than to a specified value.



Transformed beta family



Transformed/inverse transformed gamma family

Chapter 5

Calculations involving coverage modifications

5.1 Introduction

We have seen a variety of examples that involve functions of random variables. In this chapter we relate those functions to insurance applications. Throughout this chapter we assume that all random variables have support on all or a subset of the non-negative real numbers.

5.2 Deductibles

Insurance policies are often sold with a per loss deductible of d. When the loss, x, is at or below d the insurance pays nothing. When the loss is above d, the insurance pays $x - d$. In the language of Chapter 3 such a deductible can be defined as follows.

Definition 5.1 *An* ***ordinary deductible*** *modifies a random variable into either the excess loss or left censored and shifted variable (see page 17). The difference depends on whether the result of applying the deductible is to be per payment or per loss, respectively.*

This concept has already been introduced along with formulas for determining its moments. The density and distribution functions may be found as follows. For the excess loss variable,

$$f_Y(y) = \frac{f_X(y+d)}{1 - F_X(d)},\ y > 0 \tag{5.1}$$

noting that for a discrete distribution, the density function need only be replaced by the probability function. The distribution function is

$$F_Y(y) = \frac{F_X(y+d) - F_X(d)}{1 - F_X(d)},\ y > 0.$$

Note that as a per payment variable, the excess loss variable places no probability at 0.

The left censored and shifted variable always has discrete probability at zero of $F_X(d)$ representing the probability that a payment of zero is made because the loss did not exceed d. Above zero, the density and distribution functions are

$$f_Y(y) = f_X(y+d),\ F_Y(y) = F_X(y+d),\ y > 0. \tag{5.2}$$

Example 5.2 *Determine similar quantities for Model 2 for an ordinary deductible of 500.*

Using the above formulas, for the excess loss variable,

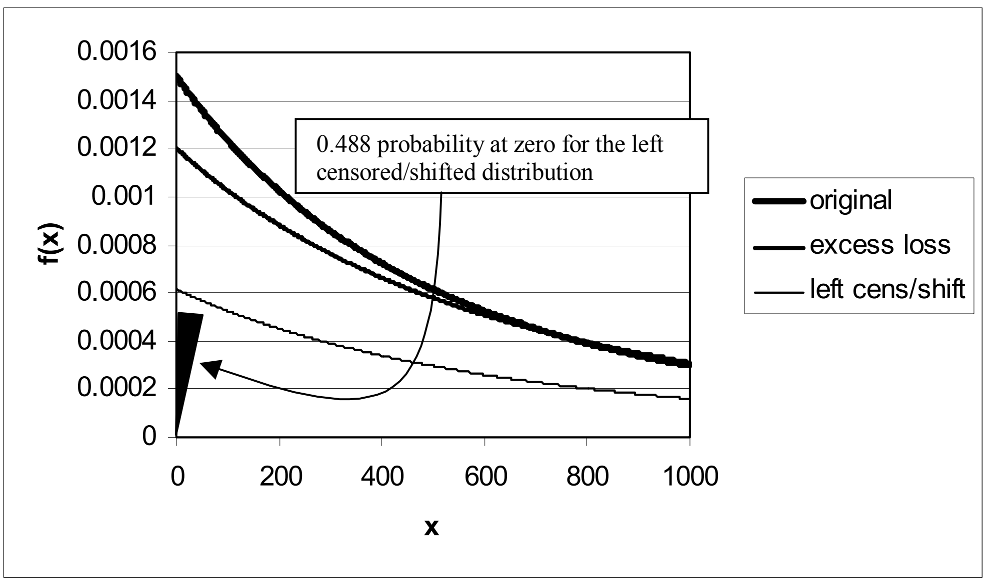
$$f_Y(y) = \frac{3(2000)^3(2000+y+500)^{-4}}{(2000)^3(2000+500)^{-3}} = \frac{3(2500)^3}{(2500+y)^4},\ F_Y(y) = 1 - \left(\frac{2500}{2500+y}\right)^3.$$

For the left censored and shifted variable,

$$f_Y(y) = \begin{cases} 0.488, & y = 0 \\ \frac{3(2000)^3}{(2500+y)^4}, & y > 0 \end{cases},\ F_Y(y) = \begin{cases} 0.488, & y = 0 \\ 1 - \frac{(2000)^3}{(2500+y)^3}, & y > 0 \end{cases}.$$

Note that the excess loss variable also has the Pareto distribution.

The figure below contains a plot of the densities. The modified densities are created as follows. For the excess loss variable, take the portion of the original density from 500 and above. Then shift it to start at zero and then multiply it by a constant so that the area under it is still one. The left censored and shifted variable also takes the original density function above 500 and shifts it to the origin, but then leaves it alone. The remaining probability is concentrated at zero, rather than spread out.



Densities for Example 5.2

□

Exercise 46 *Perform the same calculations for Model 4, this time using an ordinary deductible of 5000.*

An alternative to the ordinary deductible is the franchise deductible. This deductible differs from the ordinary deductible in that when the loss exceeds the deductible, the loss is paid in full. One example is in disability insurance where, for example, if a disability lasts seven or fewer days, no benefits are paid. However, if the disability lasts more than seven days, daily benefits are paid retroactively to the onset of the disability.

Definition 5.3 *A* ***franchise deductible*** *modifies the ordinary deductible by adding the deductible when there is a positive amount paid.*

The terms left censored and shifted and excess loss are not used here. Because this modification is unique to insurance applications, we will use per payment and per loss terminology. The per loss variable is

$$Y = \begin{cases} 0, & X \le d \\ X, & X > d \end{cases}$$

while the per payment variable is

$$Y = \begin{cases} \text{undefined}, & X \le d \\ X, & X > d. \end{cases}$$

Note that as usual, the per payment variable is a conditional random variable. The quantities derived above for the ordinary deductible are now

$$f_Y(y) = \begin{cases} F_X(d), & y = 0 \\ f_X(y), & y > d \end{cases}, \; F_Y(y) = \begin{cases} F_X(d), & 0 \le y \le d \\ F_X(y), & y > d \end{cases}$$

for the per loss variable and

$$f_Y(y) = \frac{f_X(y)}{1 - F_X(d)}, \; y > d, \; F_Y(y) = \begin{cases} 0, & 0 \le y \le d \\ \frac{F_X(y) - F_X(d)}{1 - F_X(d)}, & y > d \end{cases}$$

for the per payment variable.

Example 5.4 *Repeat Example 5.2 for a franchise deductible.*

Using the above formulas, for the per payment variable,

$$f_Y(y) = \frac{3(2000)^3(2000+y)^{-4}}{(2000)^3(2000+500)^{-3}} = \frac{3(2500)^3}{(2000+y)^4}, \; F_Y(y) = 1 - \left(\frac{2500}{2000+y}\right)^3, \; y > 500.$$

For the per loss variable,

$$f_Y(y) = \begin{cases} 0.488, & y = 0 \\ \frac{3(2000)^3}{(2000+y)^4}, & y > 500 \end{cases}, \; F_Y(y) = \begin{cases} 0.488, & 0 \le y \le 500 \\ 1 - \frac{(2000)^3}{(2000+y)^3}, & y > 500 \end{cases}.$$

□

Exercise 47 *Repeat Exercise 46 for a franchise deductible.*

Expected costs for the two types of deductible may also be calculated.

Theorem 5.5 *For an ordinary deductible, the expected cost per loss is $E(X) - E(X \wedge d)$ and the expected cost per payment is $\frac{E(X)-E(X\wedge d)}{1-F(d)}$. For a franchise deductible the expected cost per loss is $E(X) - E(X \wedge d) + d[1 - F(d)]$ and the expected cost per payment is $\frac{E(X)-E(X\wedge d)}{1-F(d)} + d$.*

Proof: For the per loss expectation with an ordinary deductible, we have, from Equations (3.7) and (3.10) that the expectation is $E(X) - E(X \wedge d)$. From Equations (5.1) and (5.2) we see that to change to a per payment basis, division by $1 - F(d)$ is required. The adjustments for the franchise deductible come from the fact that when there is a payment it will exceed that for the ordinary deductible by d. □

Example 5.6 *Determine the four expectations for Model 2 using a deductible of 500.*

Expectations could be obtained directly from the density functions obtained in Examples 5.2 and 5.4. Using Theorem 5.5 and recognizing that we have a Pareto distribution we can also look up the required values (the formulas are in *LMA*). That is,

$$F(500) = 1 - \left(\frac{2000}{2000+500}\right)^3 = 0.488, \; E(X \wedge 500) = \frac{2000}{2}\left[1 - \left(\frac{2000}{2000+500}\right)^2\right] = 360.$$

With $E(X) = 1000$ we have, for the ordinary deductible, the expected cost per loss is $1000 - 360 = 640$ while the expected cost per payment is $640/0.512 = 1250$. For the franchise deductible the expectations are $640 + 500(1 - 0.488) = 896$ and $1250 + 500 = 1750$. □

Exercise 48 *Repeat Example 5.6 for Model 4 and a 5000 deductible.*

Exercise 49 *(*) Risk 1 has a Pareto distribution with parameters $\alpha > 2$ and θ. Risk 2 has a Pareto distribution with parameters 0.8α and θ. Each risk is covered by a separate policy, each with an ordinary deductible of k. Determine the expected cost per loss for risk 1. Determine the limit as k goes to infinity of the ratio of the expected cost per loss for risk 2 to the expected cost per loss for risk 1.*

Exercise 50 *(*) Losses (prior to any deductibles being applied) have a distribution as reflected in the following table.*

x	$F(x)$	$E(X \wedge x)$
10,000	0.60	6,000
15,000	0.70	7,700
22,500	0.80	9,500
32,500	0.90	11,000
∞	1.00	20,000

There is a per loss ordinary deductible of 10,000. The deductible is then raised so that half the number of losses exceed the new deductible as exceeded the old deductible. Determine the percentage change in the expected cost per payment when the deductible is raised.

5.3 The loss elimination ratio and the effect of inflation for ordinary deductibles

A ratio that can be meaningful in evaluating the impact of a deductible is the loss elimination ratio.

Definition 5.7 *The* ***loss elimination ratio*** *is the ratio of the decrease in the expected payment with an ordinary deductible to the expected payment without the deductible.*

While many types of coverage modification can decrease the expected payment, the term loss elimination ratio is reserved for the effect of changing the deductible. Without the deductible, the expected payment is $E(X)$. With the deductible, the expected payment (from Theorem 5.5) is $E(X) - E(X \wedge d)$. Therefore the loss elimination ratio is

$$\frac{E(X) - [E(X) - E(X \wedge d)]}{E(X)} = \frac{E(X \wedge d)}{E(X)}.$$

Example 5.8 *Determine the loss elimination ratio for Model 2 with an ordinary deductible of 500.*

From Example 5.6 we have a loss elimination ratio of $360/1000 = 0.36$. Thus 36% of losses can be eliminated by introducing an ordinary deductible of 500. □

Exercise 51 *Determine the loss elimination ratio for Model 4 with an ordinary deductible of 5000.*

Inflation increases costs, but it turns out that when there is a deductible, the effect of inflation is magnified. First, some events that formerly produced losses below the deductible will now lead to payments. Second, the relative effect of inflation is magnified because the deductible is subtracted after inflation. For example, suppose an event formerly produced a loss of 600. With a 500 deductible the payment is 100. Inflation at 10% will increase the loss to 660 and the payment to 160, a 60% increase in the cost to the insurer.

Theorem 5.9 *For an ordinary deductible of* d *and uniform inflation of* $1+r$*, the expected cost per loss is* $(1+r)\{E(X) - E[X \wedge d/(1+r)]\}$*. The per payment expected cost is obtained by dividing by* $1 - F[d/(1+r)]$*.*

Proof: After inflation, losses are given by the random variable $Y = (1+r)X$. From Theorem 4.19, $f_Y(y) = f_X[y/(1+r)]/(1+r)$ and $F_Y(y) = F_X[y/(1+r)]$. Using Equation (3.8),

$$\begin{aligned}
E(Y \wedge d) &= \int_0^d y f_Y(y)dy + d[1 - F_Y(d)] \\
&= \int_0^d y f_X[y/(1+r)]dy/(1+r) + d\{1 - F_X[d/(1+r)]\} \\
&= \int_0^{d/(1+r)} (1+r)x f_X(x)dx + d\{1 - F_X[d/(1+r)]\} \\
&= (1+r)\left\{\int_0^{d/(1+r)} x f_X(x)dx + \frac{d}{1+r}[1 - F_X[d/(1+r)]]\right\} \\
&= (1+r)E[X \wedge d/(1+r)]
\end{aligned}$$

where the third line follows from the substitution $x = y/(1+r)$. Noting that $E(Y) = (1+r)E(X)$ completes the first statement of the theorem and the per payment result follows from the relationship between the distribution functions of Y and X. □

Example 5.10 *Determine the effect of inflation at 10% on an ordinary deductible of 500 applied to Model 2.*

From Example 5.6 the expected costs are 640 and 1250 per loss and per payment respectively. With 10% inflation we need

$$\begin{aligned} E(X \wedge 500/1.1) &= E(X \wedge 454.55) \\ &= \frac{2000}{2}\left[1 - \left(\frac{2000}{2000+454.55}\right)^2\right] = 336.08. \end{aligned}$$

The expected cost per loss after inflation is $1.1[1000 - 336.08] = 730.32$, an increase of 14.11%. On a per payment basis we need

$$\begin{aligned} F_Y(500) &= F_X(454.55) \\ &= 1 - \left(\frac{2000}{2000+454.55}\right)^3 \\ &= 0.459. \end{aligned}$$

The expected cost per payment is $730.32/(1 - 0.459) = 1350$, an increase of 8%. □

Exercise 52 *Determine the effect of inflation at 10% on an ordinary deductible of 5000 applied to Model 4.*

Exercise 53 *(*) Losses have a lognormal distribution with $\mu = 7$ and $\sigma = 2$. There is a deductible of 2000 and 10 losses are expected each year. Determine the loss elimination ratio. If there is uniform inflation of 20% but the deductible remains at 2000, how many payments will be expected?*

Exercise 54 *(*) Losses have a Pareto distribution with $\alpha = 2$ and $\theta = k$. There is an ordinary deductible of $2k$. Determine the loss elimination ratio before and after 100% inflation.*

Exercise 55 *(*) Losses have an exponential distribution with a mean of 1000. There is a deductible of 500. Determine the amount by which the deductible would have to be raised to double the loss elimination ratio.*

Exercise 56 *(*) The following values are available for a random variable X.*

x	$F(x)$	$E(X \wedge x)$
10,000	0.60	6,000
15,000	0.70	7,700
22,500	0.80	9,500
∞	1.00	20,000

There is a deductible of 15,000 per loss and no policy limit. Determine the expected cost per payment using X and then assuming 50% inflation (with the deductible remaining at 15,000).

Exercise 57 *(*) Losses have a lognormal distribution with $\mu = 6.9078$ and $\sigma = 1.5174$. Determine the ratio of the loss elimination ratio at 10,000 to the loss elimination ratio at 1,000. Then determine the percentage increase in the number of losses that exceed 1,000 if all losses are increased by 10%.*

Exercise 58 *(*) Losses have a mean of 2000. With a deductible of 1000 the loss elimination ratio is 0.3. The probability of a loss being greater than 1000 is 0.4. Determine the average size of a loss that is less than or equal to 1000.*

5.4 Limits

The opposite of a deductible is a policy limit. A **policy limit** is the maximum possible payment from a single claim. The typical policy limit arises in a contract where for losses below u, the insurance pays the full loss, but for losses above u the insurance pays only u. The effect of the limit is to produce a right censored random variable. It will have a mixed distribution with distribution and density function given by (where Y is random variable after the limit has been imposed)

$$F_Y(y) = \begin{cases} F_X(y), & y < u \\ 1, & y \geq u \end{cases}$$

and

$$f_Y(y) = \begin{cases} f_X(y), & y < u \\ 1 - F_X(u), & y = u. \end{cases}$$

The effect of inflation can be calculated using the following Theorem.

Theorem 5.11 *For a policy limit of u and uniform inflation of $1+r$ the expected cost is $(1+r)E[X \wedge u/(1+r)]$.*

Proof: The expected cost is $E(Y \wedge u)$. The proof of Theorem 5.9 shows that this equals the expression given in this Theorem. □

For policy limits the concept of per payment and per loss is not relevant. All losses that produced payments prior to imposing the limit will produce payments after the limit is imposed.

Example 5.12 *Impose a limit of 3000 on the distribution in Model 2. Determine the expected cost per loss with the limit as well as the proportional reduction in expected cost. Repeat these calculations after 10% uniform inflation is applied.*

For this Pareto distribution, the expected cost is

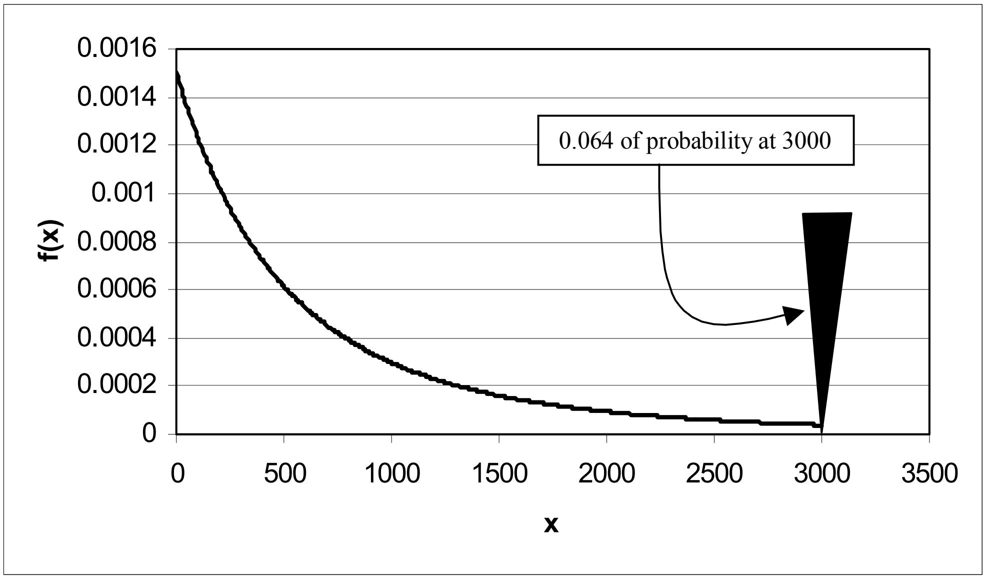
$$E(X \wedge 3000) = \frac{2000}{2}\left[1 - \left(\frac{2000}{2000+3000}\right)^2\right] = 840$$

and the proportional reduction is $(1000 - 840)/1000 = 0.16$. After inflation the expected cost is

$$1.1E(X \wedge 3000/1.1) = 1.1\frac{2000}{2}\left[1 - \left(\frac{2000}{2000+3000/1.1}\right)^2\right] = 903.11$$

for a proportional reduction of $(1100 - 903.11)/1100 = 0.179$. Also note that after inflation the expected cost has increased 7.51%, less than the general inflation rate. The effect is the opposite of the deductible—inflation is tempered, not exacerbated.

The following picture shows the density function for the right censored random variable. From 0 to 3000 it matches the original Pareto distribution. The probability of exceeding 3000, $\Pr(X > 3000) = (2000/5000)^3 = 0.064$ is concentrated at 3000.



Density function for Example 5.12

□

Exercise 59 *Determine the effect of 10% inflation on a policy limit of 150,000 on Model 4.*

A policy limit and an ordinary deductible go together in the sense that whichever applies to the insurance company's payments, the other applies to the policyholder's payments. For example, when the policy has a deductible of 500, the cost per loss to the policyholder is a random variable that is right censored at 500. When the policy has a limit of 3000, the policyholder's payments are a variable that is left truncated and shifted (as in an ordinary deductible). The opposite of the franchise deductible is a coverage that right truncates any losses (see Exercise 19 on Page 20). This coverage is rarely, if ever sold. (Would you buy a policy that pays you nothing if your loss exceeds u?)

Exercise 60 *(*) X has a Pareto distribution with $\alpha = 2$ and $\theta = 100$. Determine the range of the mean residual life function $e(d)$ as d ranges over all positive numbers. Then let $Y = 1.1X$. Determine the range of the ratio $e_Y(d)/e_X(d)$ as d ranges over all positive numbers. Finally, let Z be X right censored at 500 (that is, a limit of 500 is applied to X). Determine the range of $e_Z(d)$ as d ranges over the interval 0 to 500.*

5.5 Coinsurance

The final common coverage modification is coinsurance. In this case the insurance company pays a proportion, α, of the loss and the policyholder pays the remaining fraction. If coinsurance is the only modification, this changes the loss variable X to the payment variable, $Y = \alpha X$. The effect of multiplication has already been covered. When all four items covered in this Chapter are present (ordinary deductible, limit, coinsurance, and inflation), we create the following per loss random variable.

$$Y = \begin{cases} 0, & X < d/(1+r) \\ \alpha[(1+r)X - d], & d/(1+r) \le X < u/(1+r) \\ \alpha(u-d), & X \ge u/(1+r). \end{cases}$$

For this definition, the quantities are applied in a particular order. In particular, the coinsurance is applied last. For the contract illustrated above, the policy limit is $\alpha(u-d)$, the maximum amount payable. In this definition, u is the loss above which no additional benefits are paid and will be called the **maximum covered loss**.[1] For the per payment variable, Y is undefined for $X < d/(1+r)$.

Previous results in this Note can be combined to produce the following Theorem, given without proof.

Theorem 5.13 *For the per loss variable,*

$$E(Y) = \alpha(1+r)\{E[X \wedge u/(1+r)] - E[X \wedge d/(1+r)]\}.$$

The expected value of the per payment variable is obtained by dividing by $1 - F_X[d/(1+r)]$.

Higher moments are more difficult. The next Theorem gives the formula for the second moment. The variance can then be obtained by subtracting the square of the mean.

Theorem 5.14 *For the per loss variable*

$$E(Y^2) = \alpha^2(1+r)^2\{E[(X \wedge u^*)^2] - E[(X \wedge d^*)^2] - 2d^*E(X \wedge u^*) + 2d^*E(X \wedge d^*)\}$$

where $u^* = u/(1+r)$ *and* $d^* = d/(1+r)$. *For the second moment of the per payment variable, divide this expression by* $1 - F_X(d^*)$.

Proof: From the definition of Y,

$$E(Y^2) = \int_{d^*}^{u^*} \alpha^2[(1+r)x - d]^2 f(x)dx + \int_{u^*}^{\infty} \alpha^2(u-d)^2 f(x)dx$$

$$\begin{aligned} E(Y^2)/\alpha^2 &= (1+r)^2\left[\int_0^{u^*} x^2 f(x)dx - \int_0^{d^*} x^2 f(x)dx\right] \\ &\quad -2(1+r)d\left[\int_0^{u^*} xf(x)dx - \int_0^{d^*} xf(x)dx\right] \\ &\quad +d^2[F(u^*) - F(d^*)] + (u-d)^2[1-F(u^*)] \\ &= (1+r)^2\{E[(X \wedge u^*)^2] - u^{*2}[1-F(u^*)] - E[(X \wedge d^*)^2] + d^{*2}[1-F(d^*)]\} \\ &\quad -2(1+r)^2d^*\{E(X \wedge u^*) - u^*[1-F(u^*)] - E(X \wedge d^*) + d^*[1-F(d^*)]\} \\ &\quad +(1+r)^2d^{*2}[F(u^*) - F(d^*)] + (1+r)^2(u^* - d^*)^2[1-F(u^*)] \end{aligned}$$

$$E(Y^2)/[\alpha(1+r)]^2 = E[(X \wedge u^*)^2] - E[(X \wedge d^*)^2] - 2d^*[E(X \wedge u^*) - E(X \wedge d^*)].$$

□

[1] In the text *Loss Models* this quantity was called a policy limit (see Definition 1.7 in that text). In this Note the more conventional definition of policy limit is used and a new name (made up for this Note) given for this quantity.

Example 5.15 *Determine the mean and standard deviation per loss for Model 2 with a deductible of 500 and a policy limit of 2,500. Note that the maximum covered loss is $u = 3{,}000$.*

From earlier examples, $E(X \wedge 500) = 360$ and $E(X \wedge 3000) = 840$. The second limited moment is

$$\begin{aligned}
E[(X \wedge u)^2] &= \int_0^u x^2 \frac{3(2000)^3}{(x+2000)^4}dx + u^2\left(\frac{2000}{u+2000}\right)^3 \\
&= 3(2000)^3\int_{2000}^{u+2000}(y-2000)^2 y^{-4}dy + u^2\left(\frac{2000}{u+2000}\right)^3 \\
&= 3(2000)^3\left(-y^{-1} + 2000y^{-2} - \frac{2000^2}{3}y^{-3}\Big|_{2000}^{u+2000}\right) + u^2\left(\frac{2000}{u+2000}\right)^3 \\
&= 3(2000)^3\left[-\frac{1}{u+2000} + \frac{2000}{(u+2000)^2} - \frac{2000^2}{3(u+2000)^3}\right] \\
&\quad +3(2000)^3\left[\frac{1}{2000} - \frac{2000}{2000^2} + \frac{2000^2}{3(2000)^3}\right] + u^2\left(\frac{2000}{u+2000}\right)^3 \\
&= (2000)^2 - \left(\frac{2000}{u+2000}\right)^3(2u+2000)(u+2000).
\end{aligned}$$

Then, $E[(X \wedge 500)^2] = 160{,}000$ and $E[(X \wedge 3000)^2] = 1{,}440{,}000$ and so

$$\begin{aligned}
E(Y) &= 840 - 360 = 480 \\
E(Y^2) &= 1{,}440{,}000 - 160{,}000 - 2(500)(840) + 2(500)(360) = 800{,}000
\end{aligned}$$

for a variance of $800{,}000 - 480^2 = 569{,}600$ and a standard deviation of 754.72. □

Exercise 61 *(*) You are given that $e(0) = 25$, $S(x) = 1 - x/w$, $0 \le x \le w$, and Y is the excess loss variable for $d = 10$. Determine the variance of Y.*

Exercise 62 *(*) The loss ratio (R) is defined as total losses (L) divided by earned premiums (P). An agent will receive a bonus (B) if the loss ratio on his business is less than 0.7. The bonus is given as $B = P(0.7 - R)/3$, if this quantity is positive, otherwise it is zero. $P = 500{,}000$. L has a Pareto distribution with parameters $\alpha = 3$ and $\theta = 600{,}000$. Determine the expected value of the bonus.*

Exercise 63 *(*) Losses this year have a distribution such that $E(X \wedge d) = -0.025d^2 + 1.475d - 2.25$ for $d = 10, 11, 12, \ldots, 26$. Next year, losses will be uniformly higher by 10%. An insurance policy reimburses 100% of losses subject to a deductible of 11 up to a maximum reimbursement of 11. Determine the ratio of next year's reimbursements to this year's reimbursements.*

Exercise 64 *(*) Losses have an exponential distribution with a mean of 1000. An insurance company will pay the amount of each claim in excess of a deductible of 100. Determine the variance of the amount paid by the insurance company for one claim, including the possibility that the amount paid is zero.*

Exercise 65 *(*) Total claims for a health plan have a Pareto distribution with* $\alpha = 2$ *and* $\theta = 500$*. The health plan implements an incentive to physicians that will pay a bonus of 50% of the amount by which total claims are less than 500, otherwise no bonus is paid. It is anticipated that with the incentive plan, the claim distribution will change to become Pareto with* $\alpha = 2$ *and* $\theta = K$*. With the new distribution it turns out that expected claims plus the expected bonus is equal to expected claims prior to the bonus system. Determine the value of* K*.*

Exercise 66 *(*) In year a, total expected losses are 10,000,000. Individual losses in year a have a Pareto distribution with* $\alpha = 2$ *and* $\theta = 2000$*. A reinsurer pays the excess of each individual loss over 3000. For this, the reinsurer is paid a premium equal to 110% of expected covered losses. In year b, losses will experience 5% inflation over year a, but the frequency of losses will not change. Determine the ratio of the premium in year b to the premium in year a.*

Exercise 67 *(*) Losses have a uniform distribution from 0 to 50,000. There is a per loss deductible of 5000 and a policy limit of 20,000 (meaning that the maximum covered loss is 25,000). Determine the expected payment, given that a payment has been made.*

Exercise 68 *(*) Losses have a lognormal distribution with* $\mu = 10$ *and* $\sigma = 1$*. For losses below 50,000, no payment is made. For losses between 50,000 and 100,000, the full amount of the loss is paid. For losses in excess of 100,000 the limit of 100,000 is paid. Determine the expected cost per loss.*

Chapter 6

Three examples

6.1 Introduction

In this Chapter we present three examples. The first is a model for the time to death. The second model is for the time from when a medical malpractice incident occurs to when it is reported. The third model is for the amount of a liability payment. This model is also continuous, but most likely has a decreasing failure rate (typical of payment amount variables). On the other hand, time to event variables tend to have an increasing failure rate.

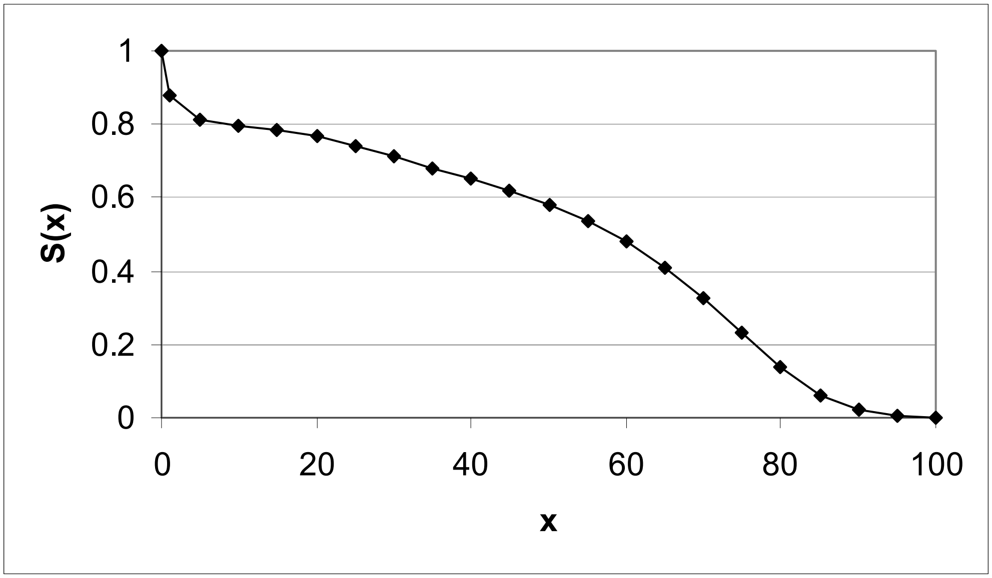
Each example begins with some data. While the numbers were made up, they are representative of the type of data one might encounter in such a situation. While the details of how one goes from data to model are postponed to Course 4, it is hard to discuss realistic models without having a context and without having some data.

6.2 Time to death

6.2.1 The data

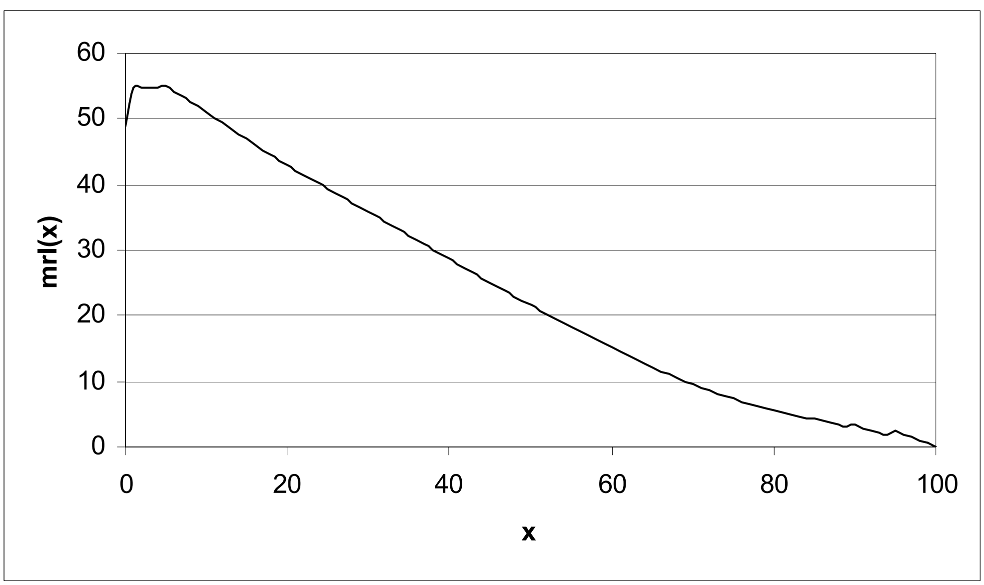
A variety of mortality tables are available from the Society of Actuaries at www.soa.org/tablemgr/tablemgr.asp. The typical mortality table provides values of the survival function at each whole number age at death. The table below represents female mortality in 1900, with only some of the data points presented. It is followed by a graph of the survival function obtained by connecting the given points with straight lines.

x	$S(x)$	x	$S(x)$	x	$S(x)$
0	1.000	35	0.681	75	0.233
1	0.880	40	0.650	80	0.140
5	0.814	45	0.617	85	0.062
10	0.796	50	0.580	90	0.020
15	0.783	55	0.534	95	0.003
20	0.766	60	0.478	100	0.000
25	0.739	65	0.410		
30	0.711	70	0.328		



Survival function for Society of Actuaries' data

The mean residual life function can be obtained by assuming that the survival function is indeed a straight line connecting each of the available points. From Equation (3.5) it can be computed as the area under the curve beyond the given age divided by the value of the survival function at that age. The following graph plots the mean residual life function.



Mean residual life function for Society of Actuaries' data

The slight increase shortly after birth indicates that in 1900 infant mortality was high. Surviving the first year after birth adds about 5 years to one's expected remaining lifetime. After that, the mean residual life steadily decreases, which is the effect of aging that we would have expected.

6.2.2 Some calculations

Items such as deductibles, limits, and coinsurances are not particularly interesting with regard to insurances on human lifetimes. We will consider the following two questions.

1. For a person age 65, determine the expected present value of providing 1000 at the beginning of each year in which the person is alive. The interest rate is 6%.
2. For a person age 20, determine the expected present value of providing 1000 at the moment of death. The interest rate is 6%.

Alert readers will recognize that these are basic calculations from Chapters 4 and 5 of *Actuarial Mathematics*. They are included here to demonstrate that many actuarial calculations are based on random variables and their moments. For the first problem, the present value random variable Y can be written as $Y = 1000(Y_0 + \cdots + Y_{34})$ where Y_j is the present value of that part of the benefit that pays 1 at age $65 + j$ if the person is alive at that time and pays nothing thereafter. Then,

$$Y_j = \begin{cases} 1.06^{-j}, & \text{with probability } S(65+j)/S(65) \\ 0, & \text{with probability } 1 - S(65+j)/S(65). \end{cases}$$

The answer is then

$$\begin{aligned} E(Y) &= 1000 \sum_{j=0}^{34} 1.06^{-j} S(65+j)/0.410 \\ &= 8408.07 \end{aligned}$$

where linear interpolation was used for intermediate values of the survival function.

For the second problem, the present value random variable is $Z = 1000(1.06^{-T})$ where T is the time in years to death of the 20-year old. The calculation is

$$E(Z) = 1000 \int_0^{80} 1.06^{-t} f(20+t)dt/S(20).$$

When linear interpolation is used to obtain the survival function at intermediate ages, the density function becomes the slope. That is, if x is a multiple of 5, then

$$f(t) = \frac{S(x) - S(x+5)}{5}, \; x < t < x+5.$$

Breaking the range of integration into 16 pieces gives

$$\begin{aligned} E(Z) &= \frac{1000}{0.766} \sum_{j=0}^{15} \frac{S(20+5j) - S(25+5j)}{5} \int_{5j}^{5+5j} 1.06^{-t} dt \\ &= \frac{200}{0.766} \sum_{j=0}^{15} [S(20+5j) - S(25+5j)] \frac{1.06^{-5j} - 1.06^{-5-5j}}{\ln 1.06} \\ &= 155.10. \end{aligned}$$

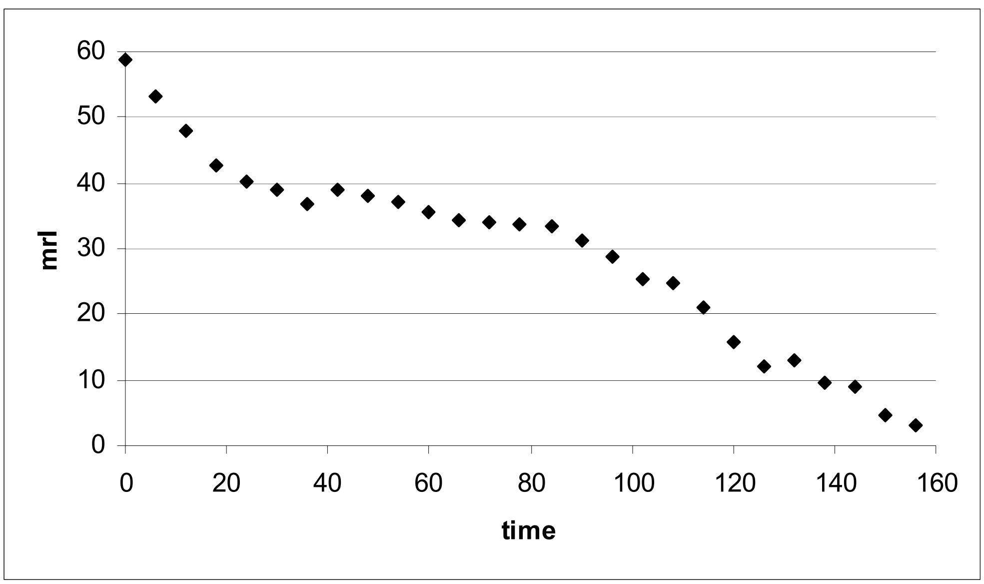
Exercise 69 *From ages 5 through 100 the mean residual life function is essentially linear. Because insurances are rarely sold under age 5, it would be reasonable to extend the graph linearly back to 0. Then a reasonable approximation is $e(x) = 60 - 0.6x$. From this, determine the density and survival function for the age at death and then use this function to solve the two problems.*

6.3 Time from incidence to report

Consider an insurance contract that provides payment when a certain event (such as death, disability, fire) occurs. There are three key dates. The first is when the event occurs, the second is when it is reported to the insurance company, and the third is when the claim is settled. The time between these dates is important because it affects the amount of interest that can be earned on the premium prior to paying the claim and because it provides a mechanism for estimating unreported claims. This example concerns the time from incidence to report. The particular example used here is found in *Loss Models* on pages 137–138 and is based on a paper by Accomando and Weissner ("Report Lag Distributions: Estimation and Application to IBNR Counts," *Transcripts of the 1988 Casualty Loss Reserve Seminar*, pp. 1038–1133.

6.3.1 The problem and some data

This example concerns medical malpractice claims that occurred in a particular year. 168 months after the beginning of the year under study, there have been 463 claims reported that were known to have occurred in that year. The distribution of the times from occurrence to report (by month in 6 month intervals) is given in Table 6.1. A graph of the mean residual life function appears below.[1]



Mean residual life function for report lag data

Your task is to fit a model to these observations and then use the model to estimate the total number of claims that occurred in the year under study. A look at the mean residual life function indicates a decreasing pattern and so a lighter-than-exponential tail is expected. A Weibull model can have such a tail and so can be used here.

[1]Because of the right truncation of the data there are some items missing for calculation of the mean residual life. It is not clear from the data what the effect will be. This picture gives a guide, but the model ultimately selected should both fit the data and be reasonable based on the analyst's experience and judgment.

Table 6.1: Medical malpractice report lags

Lag in months	No. of claims	Lag in months	No. of claims
0–6	4	84–90	11
6–12	6	90–96	9
12–18	8	96–102	7
18–24	38	102–108	13
24–30	45	108–114	5
30–36	36	114–120	2
36–42	62	120–126	7
42–48	33	126–132	17
48–54	29	132–138	5
54–60	24	138–144	8
60–66	22	144–150	2
66–72	24	150–156	6
72–78	21	156–162	2
78–84	17	162–168	0

6.3.2 Analysis

Using maximum likelihood to estimate the Weibull parameters[2] the result is $\hat{\tau} = 1.71268$ and $\hat{\theta} = 67.3002$. According to the Weibull distribution, the probability that a claim is reported as of time 168 is

$$F(168) = 1 - e^{-(168/\theta)^\tau}.$$

If N is the unknown total number of claims, the number observed by time 168 is the result of binomial sampling, and thus on an expected value basis we obtain

$$\text{Expected number of reported claims by time } 168 = N[1 - e^{-(168/\theta)^\tau}].$$

Setting this expectation equal to the observed number reported of 463 and then solving for N yields

$$N = \frac{463}{1 - e^{-(168/\theta)^\tau}}$$

Inserting the parameter estimates yields the value 466.88. Thus, after 14 years, we expect to have about 4 more claims reported.

In Course 4, material is also presented on constructing confidence intervals. An approximate 95% confidence interval is 466.88 ± 2.90, indicating that there could reasonably be between 1 and 7 additional claims reported.

6.4 Payment amount

You are the consulting actuary for a reinsurer and have been asked to determine the expected cost and the risk (as measured by the coefficient of variation) for various coverages. To help you out,

[2]Parameter estimation techniques are covered in Exam 4 along with methods for deciding that a Weibull model is appropriate for these data.

losses from 200 claims have been supplied. The reinsurer also estimates (and you may confidently rely on its estimate) that there will be 21 losses per year and the number of losses has a Poisson distribution. The coverages it is interested in are full coverage, 1 million excess of 250,000, and 2 million excess of 500,000. The phrase "z excess of y" is to be interpreted as $d = y$ and $u = y + z$ in the notation of Theorem 5.13.

6.4.1 The data and preliminary model

178 losses that were 200,000 or below (all expressed in whole numbers of thousands of dollars) that were supplied are summarized in the following table.

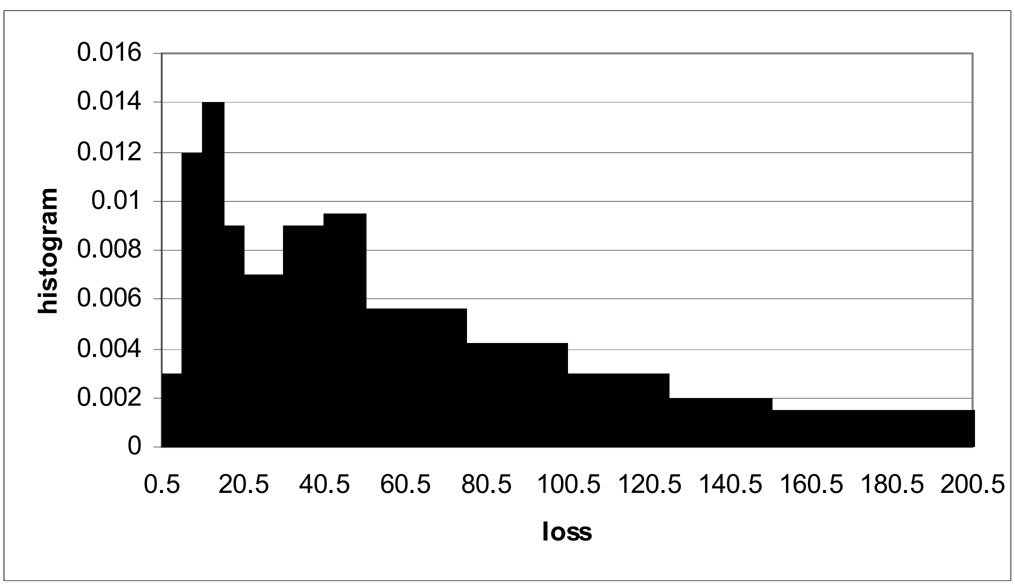
loss range	number of losses	loss range	number of losses
1–5	3	41–50	19
6–10	12	51–75	28
11–15	14	76–100	21
16–20	9	101–125	15
21–25	7	126–150	10
26–30	7	151–200	15
31–40	18		

In addition, there were 22 losses in excess of 200. They are listed below.

206	219	230	235	241	272	283	286	312	319	385
427	434	555	562	584	700	711	869	980	999	1506

Finally, the 178 losses in the table sum to 11,398 and their squares sum to 1,143,164.

To get a feel for the data in the table, the following histogram was constructed. Keep in mind that the height of a histogram bar is the count in the cell divided by the sample size (200), and then further divided by the interval width. Therefore, the first bar has a height of $\frac{3}{200(5)} = 0.003$.



Histogram of losses

It can be seen from the histogram that the underlying distribution has a non-zero mode. To check the tail, we can compute the empirical mean residual life function at a number of values. They are presented below.

loss	mean residual life
200	314
300	367
400	357
500	330
600	361
700	313
800	289
900	262

The function appears to be fairly constant and so an exponential model seems reasonable.

6.4.2 The final model

A 2-component spliced model was selected. The empirical model is used through 200 (thousand) and an exponential model thereafter. There are (at least) two ways to choose the exponential model. One is to restrict the parameter by forcing the distribution to place 11% (22 out of 200) of probability at points above 200. The other option is to estimate the exponential model independent of the 11% requirement and then multiply the density function to make the area above 200 be 0.11. The latter was selected and the resulting parameter estimate is $\theta = 314$. Estimation of parameters and model selection issues belong in Exam 4, but it seemed appropriate to give a little background here to make the example more complete. For values below 200, the empirical distribution places probability 1/200 at each observed value.

The resulting exponential density function (for $x > 200$) is

$$f(x) = 0.000662344 e^{-x/314}.$$

The graph on the next page shows the empirical mean residual life values and the constant mean residual life from the exponential model.

6.4.3 Moments

For a coverage that pays all losses, the kth moment is (where the 200 losses in the sample have been ordered from smallest to largest)

$$\begin{aligned} E(X^k) &= \frac{1}{200}\sum_{j=1}^{178} x_j^k + \int_{200}^{\infty} x^k f(x)dx \\ E(X) &= \frac{11{,}398}{200} + 0.000662344[314(200) + 314^2]e^{-200/314} = 113.53 \\ E(X^2) &= \frac{1{,}143{,}164}{200} + 0.000662344[314(200)^2 + 2(314)^2(200) + 2(314)^3]e^{-200/314} = 45{,}622.93 \end{aligned}$$

The variance is $45{,}622.93 - 113.53^2 = 32{,}733.87$ for a coefficient of variation of 1.59. However, these are for one loss only. The distribution of annual losses follows a compound Poisson distribution.

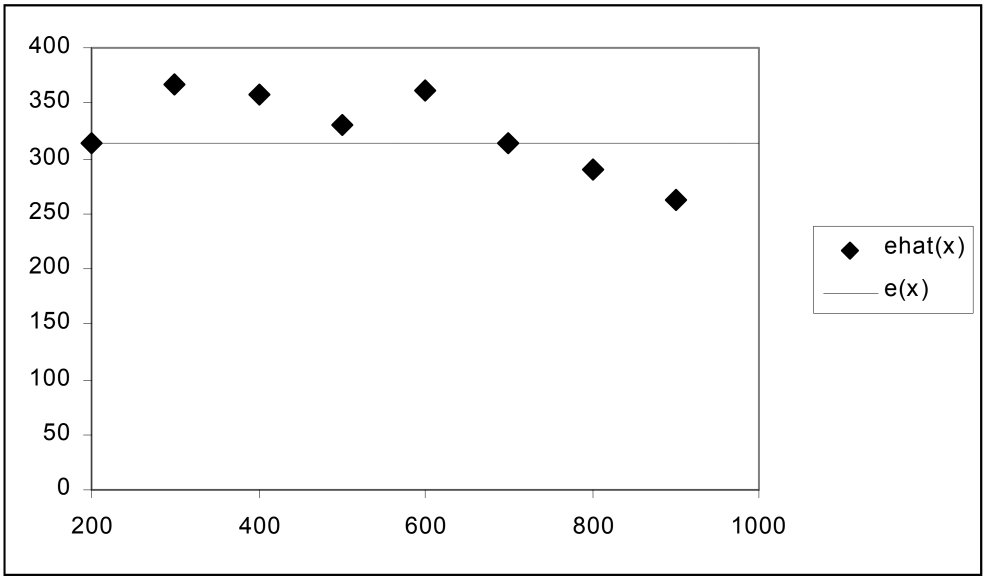
From Equation (4.6) on page 298 of *Loss Models*, the mean is

$$E(S) = E(N)E(X) = 21(113.53) = 2{,}384.13$$

and the variance is

$$\begin{aligned} Var(S) &= E(N)Var(X) + Var(N)E(X)^2 \\ &= 21(32{,}733.87) + 21(113.53)^2 = 958{,}081.53 \end{aligned}$$

for a coefficient of variation of 0.41.



Empirical and model mean residual life

For the other coverages we need general formulas for the first two limited expected moments. For $u > 200$,

$$\begin{aligned} E(X \wedge u) &= 56.99 + \int_{200}^{u} x f(x)dx + \int_{u}^{\infty} u f(x)dx \\ &= 56.99 + c\int_{200}^{u} x e^{-x/314}dx + c\int_{u}^{\infty} u e^{-x/314}dx \\ &= 56.99 + c\left(-314xe^{-x/314} - 314^2 e^{-x/314}\right)\Big|_{200}^{u} + -cu314e^{-x/314}\Big|_{u}^{\infty} \\ &= 56.99 + c\left(161{,}396e^{-200/314} - 314^2 e^{-u/314}\right) \end{aligned}$$

where $c = 0.000662344$ and similarly

$$\begin{aligned} E[(X \wedge u)^2] &= 5{,}715.82 + c\int_{200}^{u} x^2 e^{-x/314}dx + c\int_{u}^{\infty} u^2 e^{-x/314}dx \\ & \quad 5{,}715.82 + c\left[-314x^2 - 314^2(2x) - 314^3(2)\right] e^{-x/314}\Big|_{200}^{u} \\ & \quad - cu^2 314 e^{-x/314}\Big|_{u}^{\infty} \\ &= 5{,}715.82 + c\left[113{,}916{,}688e^{-200/314} - (197192u + 61{,}918{,}288)e^{-u/314}\right]. \end{aligned}$$

The following table gives the quantities needed to complete the assignment:

u	$E(X \wedge u)$	$E[(X \wedge u)^2]$
250	84.07	12,397.08
500	100.24	23,993.47
1250	112.31	41,809.37
2500	113.51	45,494.83

The requested moments for the 1000 excess of 250 coverage are, for one loss,

$$\begin{aligned}\text{Mean} &= 112.31 - 84.07 = 28.24\\ \text{Second moment} &= 41{,}809.37 - 12{,}397.08 - 2(250)(28.24) = 15{,}292.29\\ \text{Variance} &= 15{,}292.29 - 28.24^2 = 14{,}494.79\\ \text{Coefficient of variation} &= \sqrt{14{,}494.79}/28.24 = 4.26.\end{aligned}$$

It is interesting to note that while, as expected, the coverage limitations reduce the variance, the risk, as measured by the coefficient of variation has increased considerably. For a full year, the mean is 593.04, the variance is 321,138.09, and the coefficient of variation is 0.96.

For the 2000 excess of 500 coverage, we have, for one loss,

$$\begin{aligned}\text{Mean} &= 113.51 - 100.24 = 13.27\\ \text{Second moment} &= 45{,}494.83 - 23{,}993.47 - 2(500)(13.27) = 8{,}231.36\\ \text{Variance} &= 8{,}231.36 - 13.27^2 = 8{,}055.27\\ \text{Coefficient of variation} &= \sqrt{8{,}055.27}/13.27 = 6.76.\end{aligned}$$

Moving further into the tail increases our risk. For one year, the three items are 278.67, 172,858.56, and 1.49.

Appendix A

Solutions to Exercises

Exercise 1 $F_5(x) = 1 - S_5(x) = \begin{cases} 0.01x, & 0 \le x < 50 \\ 0.02x - 0.5, & 50 \le x < 75. \end{cases}$

$$f_5(x) = F_5'(x) = \begin{cases} 0.01, & 0 < x < 50 \\ 0.02, & 50 \le x < 75. \end{cases} \qquad h_5(x) = \frac{f_5(x)}{S_5(x)} = \begin{cases} \frac{1}{100-x}, & 0 < x < 50 \\ \frac{1}{75-x}, & 50 \le x < 75. \end{cases}$$

Exercise 2 The requested plots appear below. The triangular spike at zero in the density function for Model 4 indicates the 0.7 of discrete probability at zero.

Distribution function for Model 3

Distribution function for Model 4

Distribution function for Model 5

Probability function for Model 3

Density function for Model 4

Density function for Model 5

Hazard rate for Model 4

Hazard rate for Model 5

Exercise 3 $f'(x) = 4(1+x^2)^{-3} - 24x^2(1+x^2)^{-4}$. Setting the derivative equal to zero and multiplying by $(1+x^2)^4$ gives the equation $4(1+x^2) - 24x^2 = 0$. This is equivalent to $x^2 = 1/5$. The only positive solution is the mode of $1/\sqrt{5}$.

Exercise 4 The survival function can be recovered as

$$\begin{aligned} 0.5 &= S(0.4) = e^{-\int_0^{0.4} A+e^{2x}dx} \\ &= e^{-Ax-0.5e^{2x}\big|_0^{0.4}} \\ &= e^{-0.4A-0.5e^{0.8}+0.5}. \end{aligned}$$

Taking logarithms gives

$$-0.693147 = -0.4A - 1.112770 + 0.5$$

and thus $A = 0.2009$.

Exercise 5 The ratio is

$$\begin{aligned} r &= \frac{\left(\frac{10{,}000}{10{,}000+d}\right)^2}{\left(\frac{20{,}000}{20{,}000+d^2}\right)^2} \\ &= \left(\frac{20{,}000 + d^2}{20,000 + 2d}\right)^2 \\ &= \frac{20{,}000^2 + 40{,}000d^2 + d^4}{20{,}000^2 + 80{,}000d + 4d^2}. \end{aligned}$$

From observation, or two applications of L'Hôpital's rule, we see that the limit is infinity.

Exercise 6

$$\mu_3 = \int_{-\infty}^{\infty}(x-\mu)^3 f(x)dx = \int_{-\infty}^{\infty}(x^3 - 3x^2\mu + 3x\mu^2 - \mu^3)f(x)dx = \mu_3' - 3\mu_2'\mu + 2\mu^3.$$

$$\mu_4 = \int_{-\infty}^{\infty}(x-\mu)^4 f(x)dx = \int_{-\infty}^{\infty}(x^4 - 4x^3\mu + 6x^2\mu^2 - 4x\mu^3 + \mu^4)f(x)dx = \mu_4' - 4\mu_3'\mu + 6\mu_2'\mu^2 - 3\mu^4.$$

Exercise 7 For Model 1, $\sigma^2 = 3{,}333.33 - 50^2 = 833.33$, $\sigma = 28.8675$. $\mu_3' = \int_0^{100} x^3(.01)dx = 250{,}000$, $\mu_3 = 0$, $\gamma_1 = 0$. $\mu_4' = \int_0^{100} x^4(.01)dx = 20{,}000{,}000$, $\mu_4 = 1{,}250{,}000$, $\gamma_2 = 1.8$.

For Model 2, $\sigma^2 = 4{,}000{,}000 - 1000^2 = 3{,}000{,}000$, $\sigma = 1732.05$. μ_3' and μ_4' are both infinite so the skewness and kurtosis are not defined.

For Model 3, $\sigma^2 = 2.25 - .93^2 = 1.3851$, $\sigma = 1.1769$.
$\mu_3' = 0(.5) + 1(.25) + 8(.12) + 27(.08) + 64(.05) = 6.57$, $\mu_3 = 1.9012$, $\gamma_1 = 1.1663$.
$\mu_4' = 0(.5) + 1(.25) + 16(.12) + 81(.08) + 256(.05) = 21.45$, $\mu_4 = 6.4416$, $\gamma_2 = 3.3576$.

For Model 4, $\sigma^2 = 6{,}000{,}000{,}000 - 30{,}000^2 = 5{,}100{,}000{,}000$, $\sigma = 71{,}414$.
$\mu_3' = 0^3(.7) + \int_0^{\infty} x^3(.000003)e^{-.00001x}dx = 1.8 \times 10^{15}$, $\mu_3 = 1.314 \times 10^{15}$, $\gamma_1 = 3.6078$.
$\mu_4' = \int_0^{\infty} x^4(.000003)e^{-.00001x}dx = 7.2 \times 10^{20}$, $\mu_4 = 5.3397 \times 10^{20}$, $\gamma_2 = 20.5294$.

For Model 5, $\sigma^2 = 2395.83 - 43.75^2 = 481.77$, $\sigma = 21.95$.
$\mu_3' = \int_0^{50} x^3(.01)dx + \int_{50}^{75} x^3(.02)dx = 142{,}578.125$, $\mu_3 = -4394.53$, $\gamma_1 = -0.4156$.
$\mu_4' = \int_0^{50} x^4(.01)dx + \int_{50}^{75} x^4(.02)dx = 8{,}867{,}187.5$, $\mu_4 = 439{,}758.30$, $\gamma_2 = 1.8947$.

For Model 6, $\sigma^2 = 42.5 - 6.25^2 = 3.4375$, $\sigma = 1.8540$.
$\mu_3' = 310.75$, $\mu_3 = 2.15625$, $\gamma_1 = .3383$.
$\mu_4' = 2424.5$, $\mu_4 = 39.0508$, $\gamma_2 = 3.3048$.

Exercise 8 The standard deviation is the mean times the coefficient of variation, or 4 and so the variance is 16. From Equation (3.3) the second raw moment is $16 + 2^2 = 20$. The third central moment is (using Exercise 6) $136 - 3(20)(2) + 2(2)^3 = 32$. The skewness is the third central moment divided by the cube of the standard deviation, or $32/4^3 = 1/2$.

Exercise 9 For a gamma distribution the mean is $\alpha\theta$. From *LMA* the second raw moment is $\alpha(\alpha+1)\theta^2$ and so the variance is $\alpha\theta^2$. The coefficient of variation is $\sqrt{\alpha\theta^2}/\alpha\theta = \alpha^{-1/2} = 1$. Therefore $\alpha = 1$. The third raw moment is $\alpha(\alpha+1)(\alpha+2)\theta^3 = 6\theta^3$. From Exercise 6, the third central moment is $6\theta^3 - 3(2\theta^2)\theta + 2\theta^3 = 2\theta^3$ and the skewness is $2\theta^3/(\theta^2)^{3/2} = 2$.

Exercise 10 For Model 1,

$$e(d) = \frac{\int_d^{100}(1 - .01x)dx}{1 - .01d} = \frac{100 - d}{2}.$$

For Model 2,

$$e(d) = \frac{\int_d^{\infty}\left(\frac{2000}{x+2000}\right)^3 dx}{\left(\frac{2000}{d+2000}\right)^3} = \frac{2000 + d}{2}$$

For Model 3,

$$e(d) = \begin{cases} \frac{.25(1-d)+.12(2-d)+.08(3-d)+.05(4-d)}{.5} = 1.86 - d, & 0 \le d < 1 \\ \frac{.12(2-d)+.08(3-d)+.05(4-d)}{.25} = 2.72 - d, & 1 \le d < 2 \\ \frac{.08(3-d)+.05(4-d)}{.13} = 3.3846 - d, & 2 \le d < 3 \\ \frac{.05(4-d)}{.05} = 4 - d, & 3 \le d < 4. \end{cases}$$

For Model 4,

$$e(d) = \frac{\int_d^\infty .3e^{-.00001x}dx}{.3e^{-.00001d}} = 100{,}000.$$

The functions are straight lines for Models 1, 2, and 4. Model 1 has negative slope, Model 2 has positive slope, and Model 4 is horizontal.

Exercise 11 For a uniform distribution on the interval from 0 to w the density function is $f(x) = 1/w$. The mean residual life is

$$\begin{aligned} e(d) &= \frac{\int_d^w (x-d)w^{-1}dx}{\int_d^w w^{-1}dx} \\ &= \frac{\left.\frac{(x-d)^2}{2w}\right|_d^w}{\frac{w-d}{w}} \\ &= \frac{(w-d)^2}{2(w-d)} \\ &= \frac{w-d}{2}. \end{aligned}$$

The equation becomes

$$\frac{w-30}{2} = \frac{100-30}{2} + 4$$

with a solution of $w = 108$.

Exercise 12 From the definition,

$$e(\lambda) = \frac{\int_\lambda^\infty (x-\lambda)\lambda^{-1}e^{-x/\lambda}dx}{\int_\lambda^\infty \lambda^{-1}e^{-x/\lambda}dx} = \lambda.$$

Exercise 13

$$\begin{aligned} E(X) &= \int_0^\infty xf(x)dx = \int_0^d xf(x)dx + \int_d^\infty df(x)dx + \int_d^\infty (x-d)f(x)dx \\ &= \int_0^d xf(x)dx + d[1-F(d)] + e(d)S(d) = E[X \wedge d] + e(d)S(d). \end{aligned}$$

Exercise 14 For Model 1, from (3.8),

$$E[X \wedge u] = \int_0^u x(0.01)dx + u(1 - 0.01u) = u(1 - 0.005u)$$

and from (3.10),

$$E[X \wedge u] = 50 - \frac{100-u}{2}(1-0.01u) = u(1-0.005u).$$

From (3.9),

$$E[X \wedge u] = -\int_{-\infty}^{0} 0dx + \int_0^u 1 - 0.01x dx = u - 0.01u^2/2 = u(1-0.005u).$$

For Model 2, from (3.8),

$$E[X \wedge u] = \int_0^u x\frac{3(2000)^3}{(x+2000)^4}dx + u\frac{2000^3}{(2000+u)^3} = 1000\left[1 - \frac{4{,}000{,}000}{(2000+u)^2}\right]$$

and from (3.10),

$$E[X \wedge u] = 1000 - \frac{2000+u}{2}\left(\frac{2000}{2000+u}\right)^3 = 1000\left[1 - \frac{4{,}000{,}000}{(2000+u)^2}\right].$$

From (3.9),

$$E[X \wedge u] = \int_0^u \left(\frac{2000}{2000+x}\right)^3 dx = \left.\frac{-2000^3}{2(2000+x)^2}\right|_0^u = 1000\left[1 - \frac{4{,}000{,}000}{(2000+u)^2}\right].$$

For Model 3, from (3.8),

$$E[X \wedge u] = \begin{cases} 0(.5) + u(.5) = .5u, & 0 \le u < 1 \\ 0(.5) + 1(.25) + u(.25) = .25 + .25u, & 1 \le u < 2 \\ 0(.5) + 1(.25) + 2(.12) + u(.13) = .49 + .13u, & 2 \le u < 3 \\ 0(.5) + 1(.25) + 2(.12) + 3(.08) + u(.05) = .73 + .05u, & 3 \le u < 4 \end{cases}$$

and from (3.10),

$$E[X \wedge u] = \begin{cases} .93 - (1.86 - u)(.5) = .5u, & 0 \le u < 1 \\ .93 - (2.72 - u)(.25) = .25 + .25u, & 1 \le u < 2 \\ .93 - (3.3846 - u)(.13) = .49 + .13u, & 2 \le u < 3 \\ .93 - (4 - u)(.05) = .73 + .05u, & 3 \le u < 4 \end{cases}.$$

For Model 4, from (3.8),

$$\begin{aligned} E[X \wedge u] &= \int_0^u x(.000003)e^{-.00001x}dx + u(.3)e^{-.00001u} \\ &= 30{,}000[1 - e^{-.00001u}] \end{aligned}$$

and from (3.10),

$$E[X \wedge u] = 30{,}000 - 100{,}000(.3e^{-.00001u}) = 30{,}000[1 - e^{-.00001u}].$$

Exercise 15 For a discrete distribution (which all empirical distributions are), the mean residual life function is

$$e(d) = \frac{\sum_{x_j > d}(x_j - d)p(x_j)}{\sum_{x_j > d} p(x_j)}.$$

When d is equal to a possible value of X, the function cannot be continuous because there is jump in the denominator, but not in the numerator. For an exponential distribution, argue as in Exercise 12 to see that it is constant. For the Pareto distribution,

$$\begin{aligned} e(d) &= \frac{E(X) - E(X \wedge d)}{S(d)} \\ &= \frac{\frac{\theta}{\alpha-1} - \frac{\theta}{\alpha-1}\left[1 - \left(\frac{\theta}{\theta+d}\right)^{\alpha-1}\right]}{\left(\frac{\theta}{\theta+d}\right)^{\alpha}} \\ &= \frac{\theta}{\alpha-1}\frac{\theta+d}{\theta} = \frac{\theta+d}{\alpha-1} \end{aligned}$$

which is increasing in d. Only the second statement is true.

Exercise 16 Applying the formula from the solution to Exercise 15 gives

$$\frac{10{,}000 + 10{,}000}{0.5 - 1} = -40{,}000$$

which cannot be correct. Recall that the numerator of the mean residual life is $E(X) - E(X \wedge d)$. However, when $\alpha \leq 1$ the expected value is infinite and so is the mean residual life.

Exercise 17 The empirical model places probability $1/n$ at each data point. Then

$$\begin{aligned} E(X \wedge 2) &= \sum_{x_j<2} x_j(1/40) + \sum_{x_j \geq 2} 2(1/40) \\ &= (20 + 15)(1/40) + (14)(2)(1/40) \\ &= 1.575. \end{aligned}$$

Exercise 18 We have

$$\begin{aligned} E(X \wedge 7000) &= \frac{1}{2000}\left(\sum_{x_j \leq 7000} x_j + \sum_{x_j > 7000} 7000\right) \\ &= \frac{1}{2000}\left(\sum_{x_j \leq 6000} x_j + \sum_{x_j > 6000} 6000 + \sum_{6000 < x_j \leq 7000} (x_j - 6000) + \sum_{x_j > 7000} 1000\right) \\ &= E(X \wedge 6000) + [200{,}000 - 30(6000) + 270(1000)]/2000 \\ &= 1955. \end{aligned}$$

Exercise 19 The right truncated variable is defined as $Y = X$ given that $X \leq u$. When $X > u$ this variable is not defined. The kth moment is

$$E(Y^k) = \frac{\int_0^u x^k f(x)dx}{F(u)} = \frac{\sum_{x_i \leq u} x_i^k p(x_i)}{F(u)}.$$

Exercise 20 For Model 2, solve $p = 1 - \left(\frac{2000}{2000+\pi_p}\right)^3$ and so $\pi_p = 2000[(1-p)^{-1/3} - 1]$ and the requested percentiles are 519.84 and 1419.95.

For Model 4, the distribution function jumps from 0 to 0.7 at zero and so $\pi_{.5} = 0$. For percentile above 70, solve $p = 1 - 0.3e^{-0.00001\pi_p}$ and so $\pi_p = -100{,}000 \ln[(1-p)/0.3]$ and so $\pi_{.8} = 40{,}546.51$.

For Model 5, the distribution function has two specifications. From $x = 0$ to $x = 50$ it rises from 0.0 to 0.5 and so for percentiles at 50 or below, the equation to solve is $p = 0.01\pi_p$ for $\pi_p = 100p$. For $50 < x \leq 75$ the distribution function rises from 0.5 to 1.0 and so for percentiles from 50 to 100 the equation to solve is $p = 0.02\pi_p - 0.5$ for $\pi_p = 50p + 25$. The requested percentiles are 50 and 65.

Model 6 is similar to Model 3 and so the requested percentiles are 6 and 7.

Exercise 21 The two percentiles imply

$$\begin{aligned} 0.1 &= 1 - \left(\frac{\theta}{\theta + \theta - k}\right)^{\alpha} \\ 0.9 &= 1 - \left(\frac{\theta}{\theta + 5\theta - 3k}\right)^{\alpha}. \end{aligned}$$

Rearranging the equations and taking their ratio yields

$$\frac{0.9}{0.1} = \left(\frac{6\theta - 3k}{2\theta - k}\right)^{\alpha} = 3^{\alpha}.$$

Taking logarithms of both sides gives $\ln 9 = \alpha \ln 3$ for $\alpha = \ln 9 / \ln 3 = 2$.

Exercise 22 The two percentiles imply

$$\begin{aligned} 0.25 &= 1 - e^{-(1000/\theta)^{\tau}} \\ 0.75 &= 1 - e^{-(100{,}000/\theta)^{\tau}}. \end{aligned}$$

Subtracting and then taking logarithms of both sides gives

$$\begin{aligned} \ln 0.75 &= -(1000/\theta)^{\tau} \\ \ln 0.25 &= -(100{,}000/\theta)^{\tau}. \end{aligned}$$

Dividing the second equation by the first gives

$$\frac{\ln 0.25}{\ln 0.75} = 100^{\tau}.$$

Finally, taking logarithms of both sides gives $\tau \ln 100 = \ln[\ln 0.25 / \ln 0.75]$ for $\tau = 0.3415$.

Exercise 23 The sum has a gamma distribution with parameters $\alpha = 16$ and $\theta = 250$. According to *LMA*, $\Pr(S_{16} > 6000) = 1 - \Gamma(16; 6000/250) = 1 - \Gamma(16; 24)$. From the Central Limit Theorem, the sum has an approximate normal distribution with mean $\alpha\theta = 4000$ and variance $\alpha\theta^2 = 1{,}000{,}000$ for a standard deviation of 1000. The probability of exceeding 6000 is $1 - \Phi[(6000 - 4000)/1000] = 1 - \Phi(2) = 0.0228$.

Exercise 24 Arguing as in the examples,

$$\begin{aligned} F_Y(y) &= \Pr(X \leq y/c) \\ &= \Phi\left[\frac{\ln(y/c) - \mu}{\sigma}\right] \\ &= \Phi\left[\frac{\ln y - (\ln c + \mu)}{\sigma}\right] \end{aligned}$$

which indicates that Y has the lognormal distribution with parameters $\mu + \ln c$ and σ. Because no parameter was multiplied by c, there is no scale parameter. To introduce a scale parameter, define the lognormal distribution function as $F(x) = \Phi\left(\frac{\ln x - \ln \nu}{\sigma}\right)$. Note that the new parameter ν is simply e^{μ}. Then, arguing as above

$$\begin{aligned} F_Y(y) &= \Pr(X \le y/c) \\ &= \Phi\left[\frac{\ln(y/c) - \ln \nu}{\sigma}\right] \\ &= \Phi\left[\frac{\ln y - (\ln c + \ln \nu)}{\sigma}\right] \\ &= \Phi\left[\frac{\ln y - \ln c\nu}{\sigma}\right] \end{aligned}$$

demonstrating that ν is a scale parameter.

Exercise 25 The following is not the only possible set of answers to this question. Model 1 is a uniform distribution on the interval 0 to 100 with parameters 0 and 100. It is also a beta distribution with parameters $a = 1$, $b = 1$, and $\theta = 100$. Model 2 is a Pareto distribution with parameters $\alpha = 3$ and $\theta = 2000$. Model 3 would not normally be considered a parametric distribution. However, we could define a parametric discrete distribution with arbitrary probabilities at $0, 1, 2, 3,$ and 4 being the parameters. Conventional usage would not accept this as a parametric distribution. Similarly, Model 4 is not a standard parametric distribution, but we could define one as having arbitrary probability p at zero and an exponential distribution elsewhere. Model 5 could be from a parametric distribution with uniform probability from a to b and a different uniform probability from b to c. Model 6 is similar to Model 3.

Exercise 26 For this year,

$$\Pr(X > d) = 1 - F(d) = \left(\frac{\theta}{\theta + d}\right)^2.$$

For next year, because θ is a scale parameter, claims will have a Pareto distribution with parameters $\alpha = 2$ and 1.06θ. That makes the probability $\left(\frac{1.06\theta}{1.06\theta + d}\right)^2$. Then

$$\begin{aligned} r &= \lim_{d\to\infty} \left[\frac{1.06(\theta + d)}{1.06\theta + d}\right]^2 \\ &= \lim_{d\to\infty} \frac{1.1236\theta^2 + 2.2472\theta d + 1.1236 d^2}{1.1236\theta^2 + 2.12\theta d + d^2} \\ &= \lim_{d\to\infty} \frac{2.2472\theta + 2.2472 d}{2.12\theta + 2d} \\ &= \lim_{d\to\infty} \frac{2.2472}{2} = 1.1236. \end{aligned}$$

Exercise 27 The mth moment of a k-point mixture distribution is

$$\begin{aligned} E(Y^m) &= \int y^m [a_1 f_{X_1}(y) + \cdots + a_k f_{X_k}(y)] dy \\ &= a_1 E(Y_1^m) + \cdots + a_k E(Y_k^m). \end{aligned}$$

For this problem, the first moment is

$$a\frac{\theta_1}{\alpha-1}+(1-a)\frac{\theta_2}{\alpha+1},\ \text{if } \alpha>1.$$

Similarly, the second moment is

$$a\frac{2\theta_1^2}{(\alpha-1)(\alpha-2)}+(1-a)\frac{2\theta_2^2}{(\alpha+1)\alpha},\ \text{if } \alpha>2.$$

Exercise 28 Using the results from Exercise 27, $E(X)=\sum_{i=1}^K a_i\mu_i'$ and for the gamma distribution this becomes $\sum_{i=1}^K a_i\alpha_i\theta_i$. Similarly, for the second moment we have $E(X^2)=\sum_{i=1}^K a_i\mu_2'$ which, for the gamma distribution, becomes $\sum_{i=1}^K a_i\alpha_i(\alpha_i+1)\theta_i^2$.

Exercise 29 Parametric distribution families: It would be difficult to consider the model in Model 1 as being from a parametric family (although the uniform distribution could be considered as a special case of the beta distribution). Model 2 is a Pareto distribution and as such is a member of the transformed beta family. As a stretch, Models 3 and 6 could be considered members of a family that places probability (the parameters) on a given number of non-negative integers. Model 4 could be a member of the "exponential plus family" where the plus means the possibility of discrete probability at zero. Creating a family for model 5 seems difficult.

Variable-component mixture distributions: Only model 5 seems to be a good candidate. It is a mixture of uniform distributions in which the component uniform distributions are on adjoining intervals.

Exercise 30 For this mixture distribution,

$$\begin{aligned}
F(5000) &= 0.75\Phi\left(\frac{5000-3000}{1000}\right)+0.25\Phi\left(\frac{5000-4000}{1000}\right)\\
&= 0.75\Phi(2)+0.25\Phi(1)\\
&= 0.75(0.9772)+0.25(0.8413)=0.9432.
\end{aligned}$$

The probability of exceeding 5000 is $1-0.9432=0.0568$.

Exercise 31 The distribution function of Z is

$$\begin{aligned}
F(z) &= 0.5\left[1-\frac{1}{1+(z/\sqrt{1000})^2}\right]+0.5\left[1-\frac{1}{1+z/1000}\right]\\
&= 1-0.5\frac{1000}{1000+z^2}-0.5\frac{1000}{1000+z}\\
&= 1-\frac{0.5(1000^2+1000z+1000^2+1000z^2)}{(1000+z^2)(1000+z)}.
\end{aligned}$$

The median is the solution to $0.5=F(m)$ or

$$\begin{aligned}
(1000+m^2)(1000+m) &= 2(1000)^2+1000m+1000m^2\\
1000^2+1000m^2+1000m+m^3 &= 2(1000)^2+1000m+1000m^2\\
m^3 &= 1000^2\\
m &= 100.
\end{aligned}$$

The distribution function of W is

$$\begin{aligned} F_W(w) &= \Pr(W \le w) = \Pr(1.1Z \le w) = \Pr(Z \le w/1.1) = F_Z(w/1.1) \\ &= 0.5\left[1 - \frac{1}{1 + (w/1.1\sqrt{1000})^2}\right] + 0.5\left[1 - \frac{1}{1 + z/1100}\right]. \end{aligned}$$

This is a 50/50 mixture of a Burr distribution with parameters $\alpha = 1$, $\gamma = 2$, and $\theta = 1.1\sqrt{1000}$ and a Pareto distribution with parameters $\alpha = 1$ and $\theta = 1100$.

Exercise 32 Right censoring creates a mixed distribution with discrete probability at the censoring point. Therefore, Z is matched with 3. X is similar to Model 5 on Page 11 which has a continuous distribution function but the density function has a jump at 2. Therefore, X is matched with 2. The sum of two continuous random variables will be continuous as well, in this case over the interval from 0 to 5. Therefore, Y is matched with 1.

Exercise 33 The density function is the sum of six functions. They are (where it is understood that the function is zero where not defined),

$$\begin{aligned} f_1(x) &= 0.03125,\ 1 \le x \le 5 \\ f_2(x) &= 0.03125,\ 3 \le x \le 7 \\ f_3(x) &= 0.09375,\ 4 \le x \le 8 \\ f_4(x) &= 0.06250,\ 5 \le x \le 9 \\ f_5(x) &= 0.03125,\ 8 \le x \le 12. \end{aligned}$$

Adding the functions yields

$$f(x) = \begin{cases} 0.03125, & 1 \le x < 3 \\ 0.06250, & 3 \le x < 4 \\ 0.15625, & 4 \le x < 5 \\ 0.18750, & 5 \le x < 7 \\ 0.15625, & 7 \le x < 8 \\ 0.09375, & 8 \le x < 9 \\ 0.03125, & 9 \le x < 12. \end{cases}$$

This is a mixture of seven uniform distributions, each being uniform over the indicated interval. The weight for mixing is the value of the density function multiplied by the width of the interval.

Exercise 34 From *LMA* we see that while the Weibull distribution has all positive moments, for the inverse Weibull moments exist only for $k < \tau$. Thus by this criterion, the inverse Weibull distribution has a heavier tail. With regard to the ratio of density functions, it is (with the inverse Weibull in the numerator and marking its parameters with asterisks)

$$\frac{\tau^*\theta^{*\tau^*}x^{-\tau^*-1}e^{-(\theta^*/x)^{\tau^*}}}{\tau\theta^{-\tau}x^{\tau-1}e^{-(x/\theta)^{\tau}}} \propto x^{-\tau-\tau^*}e^{-(\theta^*/x)^{\tau^*}+(x/\theta)^{\tau}}.$$

The logarithm is

$$(x/\theta)^{\tau} - (\theta^*/x)^{\tau^*} - (\tau + \tau^*)\ln x.$$

The middle term goes to zero, so the issue is the limit of $(x/\theta)^\tau - (\tau + \tau^*)\ln x$ which is clearly infinite. With regard to the hazard rate, for the Weibull distribution we have

$$h(x) = \frac{\tau x^{\tau-1}\theta^{-\tau}e^{-(x/\theta)^\tau}}{e^{-(x/\theta)^\tau}} = \tau x^{\tau-1}\theta^{-\tau}$$

which is clearly increasing when $\tau > 1$, constant when $\tau = 1$, and decreasing when $\tau < 1$. For the inverse Weibull,

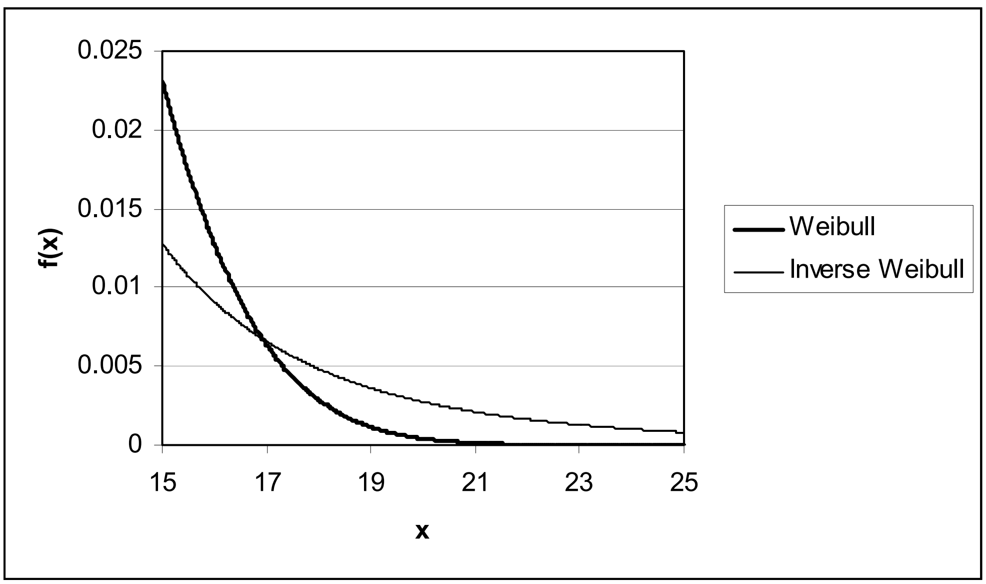
$$h(x) = \frac{\tau x^{-\tau-1}\theta^\tau e^{-(\theta/x)^\tau}}{1 - e^{-(\theta/x)^\tau}} \propto \frac{1}{x^{\tau+1}[e^{(\theta/x)^\tau} - 1]}.$$

The derivative of the denominator is

$$(\tau+1)x^\tau[e^{(\theta/x)^\tau} - 1] + x^{\tau+1}e^{(\theta/x)^\tau}\theta^\tau(-\tau)x^{-\tau-1}$$

and the limiting value of this expression is $\theta^\tau > 0$. Therefore, in the limit, the denominator is increasing and thus the hazard rate is decreasing.

The following graph displays a portion of the density function for Weibull ($\tau = 3$, $\theta = 10$) and inverse Weibull ($\tau = 4.4744$, $\theta = 7.4934$) distributions with the same mean and variance. The heavier tail of the inverse Weibull distribution is clear.



Tails of a Weibull and inverse Weibull distribution

Exercise 35 $F_Y(y) = 1 - (1 + y/\theta)^{-\alpha} = 1 - \left(\frac{\theta}{\theta+y}\right)^\alpha$. This is the cdf of the Pareto distribution. $f_Y(y) = dF_Y(y)/dy = \frac{\alpha\theta^\alpha}{(\theta+y)^{\alpha+1}}$.

Exercise 36 After three years, values are inflated by 1.331. Let X be the 1995 variable and $Y = 1.331X$ be the 1998 variable. We want

$$\Pr(Y > 500) = \Pr(X > 500/1.331) = \Pr(X > 376).$$

From the given information we have $\Pr(X > 350) = 0.55$ and $\Pr(X > 400) = 0.50$. Therefore, the desired probability must be between these two values.

Exercise 37 Inverse: $F_Y(y) = 1 - \left[1 - \left(\frac{\theta}{\theta+y^{-1}}\right)^\alpha\right] = \left(\frac{y}{y+\theta^{-1}}\right)^\alpha$. From *LMA* this is the inverse Pareto distribution with $\tau = \alpha$ and $\theta = 1/\theta$. Transformed: $F_Y(y) = 1 - \left(\frac{\theta}{\theta+y^\tau}\right)^\alpha$. This is the Burr distribution with $\alpha = \alpha$, $\gamma = \tau$, and $\theta = \theta^{1/\tau}$. Inverse transformed: $F_Y(y) = 1 - \left[1 - \left(\frac{\theta}{\theta+y^{-\tau}}\right)^\alpha\right] = \left[\frac{y^\tau}{y^\tau+(\theta^{-1/\tau})^\tau}\right]^\alpha$. This is the inverse Burr distribution with $\tau = \alpha$, $\gamma = \tau$, and $\theta = \theta^{-1/\tau}$.

Exercise 38 $F_Y(y) = 1 - \frac{(y^{-1}/\theta)^\gamma}{1+(y^{-1}/\theta)^\gamma} = \frac{1}{1+(y^{-1}/\theta)^\gamma} = \frac{(y\theta)^\gamma}{1+(y\theta)^\gamma}$. This is the loglogistic distribution with γ unchanged and $\theta = 1/\theta$.

Exercise 39 $F_Z(z) = \Phi\left[\frac{\ln(z/\theta)-\mu}{\sigma}\right] = \Phi\left[\frac{\ln z-\ln\theta-\mu}{\sigma}\right]$ which is the cdf of another lognormal distribution with $\mu = \ln\theta + \mu$, and $\sigma = \sigma$.

Exercise 40 The distribution function of Y is

$$\begin{aligned}
F_Y(y) &= \Pr(Y \le y) = \Pr[\ln(1+X/\theta) \le y] \\
&= \Pr(1 + X/\theta \le e^y) \\
&= \Pr[X \le \theta(e^y - 1)] \\
&= 1 - \left[\frac{\theta}{\theta + \theta(e^y - 1)}\right]^\alpha \\
&= 1 - \left(\frac{1}{e^y}\right)^\alpha = 1 - e^{-\alpha y}.
\end{aligned}$$

This is the distribution function of an exponential random variable with parameter $1/\alpha$.

Exercise 41 $X|\Theta = \theta$ has pdf

$$f_{X|\Theta}(x|\theta) = \frac{\tau\left[(x/\theta)^\tau\right]^\alpha \exp[-(x/\theta)^\tau]}{x\Gamma(\alpha)}$$

and Θ has pdf

$$f_\Theta(\theta) = \frac{\tau\left[(\delta/\theta)^\tau\right]^\beta \exp[-(\delta/\theta)^\tau]}{\theta\Gamma(\beta)}.$$

The mixture distribution has pdf

$$\begin{aligned}
f(x) &= \int_0^\infty \frac{\tau\left[(x/\theta)^\tau\right]^\alpha \exp[-(x/\theta)^\tau]}{x\Gamma(\alpha)} \frac{\tau\left[(\delta/\theta)^\tau\right]^\beta \exp[-(\delta/\theta)^\tau]}{\theta\Gamma(\beta)} d\theta \\
&= \frac{\tau^2 x^{\tau\alpha}\delta^{\tau\beta}}{x\Gamma(\alpha)\Gamma(\beta)} \int_0^\infty \theta^{-\tau\alpha-\tau\beta-1} \exp[-\theta^{-\tau}(x^\tau + \delta^\tau)] d\theta \\
&= \frac{\tau^2 x^{\tau\alpha}\delta^{\tau\beta}}{x\Gamma(\alpha)\Gamma(\beta)} \frac{\Gamma(\alpha+\beta)}{\tau(x^\tau+\delta^\tau)^{\alpha+\beta}} = \frac{\Gamma(\alpha+\beta)\tau x^{\tau\alpha-1}\delta^{\tau\beta}}{\Gamma(\alpha)\Gamma(\beta)(x^\tau+\delta^\tau)^{\alpha+\beta}}
\end{aligned}$$

which is a transformed beta pdf with (using the parameterization in *LMA*) $\gamma = \tau$, $\tau = \alpha$, $\alpha = \beta$, and $\theta = \delta$.

Exercise 42 The requested gamma distribution has $\alpha\theta = 1$ and $\alpha\theta^2 = 2$ for $\alpha = 0.5$ and $\theta = 2$. Then

$$\begin{aligned}
\Pr(N=1) &= \int_0^\infty \frac{e^{-\lambda}\lambda^1}{1!}\frac{\lambda^{-0.5}e^{-0.5\lambda}}{2^{0.5}\Gamma(0.5)} \\
&= \frac{1}{\Gamma(0.5)\sqrt{2}}\int_0^\infty \lambda^{0.5}e^{-1.5\lambda}d\lambda \\
&= \frac{1}{1.5\Gamma(0.5)\sqrt{3}}\int_0^\infty y^{0.5}e^{-y}dy \\
&= \frac{\Gamma(1.5)}{1.5\Gamma(0.5)\sqrt{3}} \\
&= \frac{0.5}{1.5\sqrt{3}} = 0.19245.
\end{aligned}$$

Line three follows from the substitution $y = 1.5\lambda$. Line five follows from the gamma function identity $\Gamma(1.5) = 0.5\Gamma(0.5)$. After reading Chapter 3 of *Loss Models*, it can be known in advance that N has a negative binomial distribution and its parameters can be determined by matching moments. In particular, we have $E(N) = E[E(N|\Lambda)] = E(\Lambda) = 1$ and $Var(N) = E[Var(N|\Lambda)] + Var[E(N|\Lambda)] = E(\Lambda) + Var(\Lambda) = 1 + 2 = 3$.

Exercise 43 Using the parametrization in *LMA*, the hazard rate for an exponential distribution is $h(x) = f(x)/S(x) = \theta^{-1}$. Here θ is the parameter of the exponential distribution, not the value from the exercise. But this means that the θ in the exercise is the reciprocal of the exponential parameter and thus the density function is to be written $F(x) = 1 - e^{-\theta x}$. The unconditional distribution function is

$$\begin{aligned}
F_X(x) &= \int_1^{11}(1-e^{-\theta x})0.1d\theta \\
&= 0.1(\theta + x^{-1}e^{-\theta x})\Big|_1^{11} \\
&= 1 + \frac{1}{10x}(e^{-11x} - e^{-x}).
\end{aligned}$$

Then, $S_X(0.5) = 1 - F_X(0.5) = -\frac{1}{10(0.5)}(e^{-5.5} - e^{-0.5}) = 0.1205$.

Exercise 44 We have

$$\begin{aligned}
\Pr(N \ge 2) &= 1 - F_N(1) \\
&= 1 - \int_0^5 (e^{-\lambda} + \lambda e^{-\lambda})0.2d\lambda \\
&= 1 - [-(1+\lambda)e^{-\lambda} - e^{-\lambda}]0.2\Big|_0^5 \\
&= 1 + 1.2e^{-5} + 0.2e^{-5} - 0.2 - 0.2 \\
&= 0.6094.
\end{aligned}$$

Exercise 45 Using the first definition of a spliced model we have

$$f_X(x) = \begin{cases} \tau, & 0 < x < 1000 \\ \gamma e^{-x/\theta}, & x > 1000 \end{cases}$$

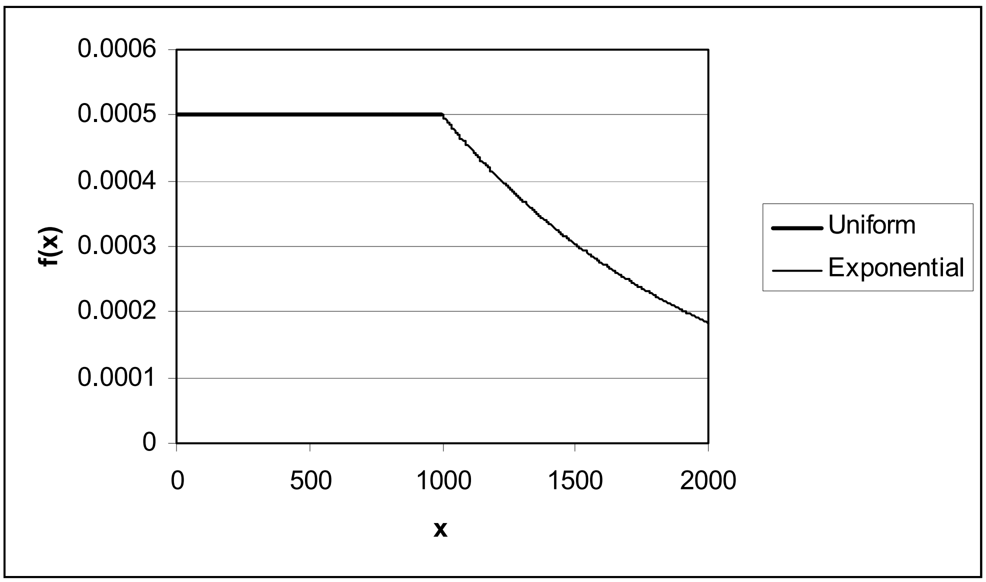
where the coefficient τ is a_1 multiplied by the the uniform density of .001 and the coefficient γ is a_2 multiplied by the scaled exponential coefficient. To ensure continuity we must have $\tau = \gamma e^{-1000/\theta}$. Finally, to ensure that the density integrates to 1, we have

$$\begin{aligned} 1 &= \int_0^{1000} \gamma e^{-1000/\theta} dx + \int_{1000}^{\infty} \gamma e^{-x/\theta} dx \\ &= 1000\gamma e^{-1000/\theta} + \gamma\theta e^{-1000/\theta} \end{aligned}$$

which implies $\gamma = [(1000+\theta)e^{-1000/\theta}]^{-1}$. The final density, a one parameter distribution, is

$$f_X(x) = \begin{cases} \frac{1}{1000+\theta}, & 0 < x \leq 1000 \\ \frac{e^{-x/\theta}}{(1000+\theta)e^{-1000/\theta}}, & x \geq 1000 \end{cases}.$$

The following graph presents this density function for the value $\theta = 1000$.



Continuous spliced density function

Exercise 46 For the excess loss variable,

$$f_Y(y) = \frac{0.000003e^{-0.00001(y+5000)}}{0.3e^{-.00001(5000)}} = 0.00001e^{-0.00001y},\ F_Y(y) = 1 - e^{-0.00001y}.$$

For the left censored and shifted variable,

$$f_Y(y) = \begin{cases} 1 - .3e^{-0.05} = 0.71463, & y = 0 \\ 0.000003e^{-0.00001(y+5000)}, & y > 0 \end{cases},\ F_Y(y) = \begin{cases} 0.71463, & y = 0 \\ 1 - 0.3e^{-0.00001(y+5000)}, & y > 0 \end{cases}$$

and it is interesting to note that the excess loss variable has an exponential distribution.

Exercise 47 For the per payment variable,

$$f_Y(y) = \frac{0.000003e^{-0.00001y}}{0.3e^{-.00001(5000)}} = 0.00001e^{-0.00001(y-5000)},\ F_Y(y) = 1 - e^{-0.00001(y-5000)},\ y > 5000.$$

For the per loss variable,

$$f_Y(y) = \begin{cases} 1 - .3e^{-0.05} = 0.71463, & y = 0 \\ 0.000003e^{-0.00001y}, & y > 5000 \end{cases}, \; F_Y(y) = \begin{cases} 0.71463, & 0 \le y \le 5000 \\ 1 - 0.3e^{-0.00001y}, & y > 5000. \end{cases}$$

Exercise 48 From Example 3.2 $E(X) = 30{,}000$ and from Exercise 14,

$$E(X \wedge 5000) = 30{,}000[1 - e^{-0.00001(5000)}] = 1463.12.$$

Also $F(5000) = 1 - 0.3e^{-0.00001(5000)} = 0.71463$ and so for an ordinary deductible the expected cost per loss is 28,536.88 and per payment is 100,000. For the franchise deductible the expected costs are 28,536.88 + 5,000(0.28537) = 29,963.73 per loss and 100,000 + 5,000 = 105,000 per payment.

Exercise 49 For risk 1,

$$\begin{aligned} E(X) - E(X \wedge d) &= \frac{\theta}{\alpha - 1} - \frac{\theta}{\alpha - 1}\left[1 - \left(\frac{\theta}{\theta + k}\right)^{\alpha - 1}\right] \\ &= \frac{\theta^\alpha}{(\alpha - 1)(\theta + k)^{\alpha - 1}}. \end{aligned}$$

The ratio is then

$$\frac{\theta^{0.8\alpha}}{(0.8\alpha - 1)(\theta + k)^{0.8\alpha - 1}} \Big/ \frac{\theta^\alpha}{(\alpha - 1)(\theta + k)^{\alpha - 1}} = \frac{(\theta + k)^{0.2\alpha}(\alpha - 1)}{\theta^{0.2\alpha}(0.8\alpha - 1)}.$$

As k goes to infinity, the limit is infinity.

Exercise 50 The expected cost per payment with the 10,000 deductible is

$$\frac{E(X) - E(X \wedge 10{,}000)}{1 - F(10{,}000)} = \frac{20{,}000 - 6{,}000}{1 - 0.60} = 35{,}000.$$

At the old deductible, 40% of losses become payments. The new deductible must have 20% of losses become payments and so the new deductible is 22,500. The expected cost per payment is

$$\frac{E(X) - E(X \wedge 22{,}500)}{1 - F(22{,}500)} = \frac{20{,}000 - 9{,}500}{1 - 0.80} = 52{,}500.$$

The increase is 17,500/35,000 = 50%.

Exercise 51 From Exercise 48, the loss elimination ratio is (30,000 − 28,536.88)/30,000 = 0.0488.

Exercise 52 With inflation at 10% we need

$$E(X \wedge 5000/1.1) = 30000[1 - e^{-0.00001(5000/1.1)}] = 1333.11.$$

After inflation, the expected cost per loss is 1.1(30,000−1333.11) = 31,533.58 an increase of 10.50%. For the per payment calculation we need $F(5000/1.1) = 1 - 0.3e^{-0.00001(5000/1.1)} = 0.71333$ for an expected cost of 110,000, an increase of exactly 10%.

Exercise 53 From *LMA*, $E(X) = \exp(7 + 2^2/2) = \exp(9) = 8103.08$. The limited expected value is

$$\begin{aligned} E(X \wedge 2000) &= e^9 \Phi\left(\frac{\ln 2000 - 7 - 2^2}{2}\right) + 2000\left[1 - \Phi\left(\frac{\ln 2000 - 7}{2}\right)\right] \\ &= 8103.08\Phi(-1.7) + 2000[1 - \Phi(0.3)] \\ &= 8103.08(0.0446) + 2000(0.3821) = 1125.60. \end{aligned}$$

The loss elimination ratio is $1125.60/8103.08 = 0.139$. With 20% inflation, the probability of exceeding the deductible is

$$\begin{aligned} \Pr(1.2X > 2000) &= \Pr(X > 2000/1.2) \\ &= 1 - \Phi\left(\frac{\ln(2000/1.2) - 7}{2}\right) \\ &= 1 - \Phi(0.2093) \\ &= 0.4171 \end{aligned}$$

and therefore, 4.171 losses can be expected to produce payments.

Exercise 54 The loss elimination ratio prior to inflation is

$$\begin{aligned} \frac{E(X \wedge 2k)}{E(X)} &= \frac{\frac{k}{2-1}\left[1 - \left(\frac{k}{2k+k}\right)^{2-1}\right]}{\frac{k}{2-1}} \\ &= \frac{2k}{2k+k} = \frac{2}{3}. \end{aligned}$$

Because θ is a scale parameter, inflation of 100% will double it to equal $2k$. Repeating the above gives the new loss elimination ratio of

$$1 - \left(\frac{2k}{2k + 2k}\right) = \frac{1}{2}.$$

Exercise 55 The original loss elimination ratio is

$$\frac{E(X \wedge 500)}{E(X)} = \frac{1000(1 - e^{-500/1000})}{1000} = 0.39347.$$

Doubling it produces the equation

$$0.78694 = \frac{1000(1 - e^{-d/1000})}{1000} = 1 - e^{-d/1000}.$$

The solution is $d = 1546$.

Exercise 56 For the current year the expected cost per payment is

$$\frac{E(X) - E(X \wedge 15{,}000)}{1 - F(15{,}000)} = \frac{20{,}000 - 7{,}700}{1 - 0.70} = 41{,}000.$$

After 50% inflation it is

$$\begin{aligned}\frac{1.5[E(X) - E(X \wedge 15{,}000/1.5)]}{1 - F(15{,}000/1.5)} &= \frac{1.5[E(X) - E(X \wedge 10{,}000)]}{1 - F(10{,}000)} \\ &= \frac{1.5(20{,}000 - 6{,}000)}{1 - 0.60} = 52{,}500.\end{aligned}$$

Exercise 57 The ratio desired is

$$\begin{aligned}\frac{E(X \wedge 10{,}000)}{E(X \wedge 1{,}000)} &= \frac{e^{6.9078+1.5174^2/2}\Phi\left(\frac{\ln 10{,}000-6.9078-1.5174^2}{1.5174}\right) + 10{,}000\left[1 - \Phi\left(\frac{\ln 10{,}000-6.9078}{1.5174}\right)\right]}{e^{6.9078+1.5174^2/2}\Phi\left(\frac{\ln 1{,}000-6.9078-1.5174^2}{1.5174}\right) + 1{,}000\left[1 - \Phi\left(\frac{\ln 1{,}000-6.9078}{1.5174}\right)\right]} \\ &= \frac{e^{8.059}\Phi(0) + 10{,}000[1 - \Phi(1.5174)]}{e^{8.059}\Phi(-1.5174) + 1{,}000[1 - \Phi(0)]} \\ &= \frac{3162(0.5) + 10{,}000(0.0647)}{3162(0.0647) + 1{,}000(0.5)} = 3.162.\end{aligned}$$

This year, the probability of exceeding 1000 is $\Pr(X > 1000) = 1 - \Phi\left(\frac{\ln 1{,}000-6.9078}{1.5174}\right) = 0.5$. With 10% inflation the distribution is lognormal with parameters $\mu = 6.9078 + \ln 1.1 = 7.0031$ and $\mu = 1.5174$. The probability is $1 - \Phi\left(\frac{\ln 1{,}000-7.0031}{1.5174}\right) = 1 - \Phi(-0.0628) = 0.525$, an increase of 5%. Alternatively, the original lognormal distribution could be used and then $\Pr(X > 1000/1.1)$ computed.

Exercise 58 The desired quantity is the expected value of a right truncated variable. It is

$$\frac{\int_0^{1000} xf(x)dx}{F(1000)} = \frac{E(X \wedge 1000) - 1000[1 - F(1000)]}{F(1000)} = \frac{E(X \wedge 1000) - 400}{0.6}.$$

From the loss elimination ratio,

$$0.3 = \frac{E(X \wedge 1000)}{E(X)} = \frac{E(X \wedge 1000)}{2000}$$

and so $E(X \wedge 1000) = 600$ making the answer $200/0.6 = 333$.

Exercise 59 From Exercise 14 we have

$$E(X \wedge 150{,}000) = 30{,}000[1 - e^{-0.00001(150{,}000)}] = 23{,}306.10.$$

After 10% inflation the expected cost is

$$1.1E(X \wedge 150{,}000/1.1) = 33{,}000[1 - e^{-0.00001(150{,}000/1.1)}] = 24{,}560.94$$

for an increase of 5.38%.

Exercise 60 From Exercise 15 we have

$$e_X(d) = \frac{\theta + d}{\alpha - 1} = \frac{100 + d}{2 - 1} = 100 + d.$$

Therefore, the range is 100 to infinity. With 10% inflation, θ becomes 110 and the mean residual life is $e_Y(d) = 110 + d$. The ratio is $\frac{110+d}{100+d}$. As d increases, the ratio decreases from 1.1 to 1. The last one has not been encountered before. The mean residual life is

$$\begin{aligned} e_Z(d) &= \frac{\int_d^{500}(x-d)f(x)dx + \int_{500}^{\infty}(500-d)f(x)dx}{1-F_X(d)} \\ &= 100 + d - \frac{(100+d)^2}{600}. \end{aligned}$$

This is a quadratic function of d. It starts at 83.33, increases to a maximum of 150 at $d = 200$, and decreases to 0 when $d = 500$. The range is 0 to 150.

Exercise 61 $25 = E(X) = \int_0^w (1 - x/w)dx = w/2$ for $w = 50$. $S(10) = 1 - 10/50 = 0.8$. Then

$$\begin{aligned} E(Y) &= \int_{10}^{50}(x-10)(1/50)dx/0.8 = 20 \\ E(Y^2) &= \int_{10}^{50}(x-10)^2(1/50)dx/0.8 = 533.33 \\ Var(Y) &= 533.33 - 20^2 = 133.33. \end{aligned}$$

Exercise 62 The bonus is $B = 500{,}000(0.7 - L/500{,}000)/3 = (350{,}000 - L)/3$, if positive. The bonus is positive provided $L < 350{,}000$. The expected bonus is

$$\begin{aligned} E(B) &= \frac{1}{3}\int_0^{350{,}000}(350{,}000 - l)f_L(l)dl \\ &= \frac{1}{3}\{350{,}000F_L(350{,}000) - E(L \wedge 350{,}000) + 350{,}000[1 - F_L(350{,}000)]\} \\ &= \frac{1}{3}\left\{350{,}000 - \frac{600{,}000}{2}\left[1 - \left(\frac{600{,}000}{950{,}000}\right)^2\right]\right\} \\ &= 56{,}556. \end{aligned}$$

Exercise 63 The quantity we seek is

$$\frac{1.1[E(X \wedge 22/1.1) - E(X \wedge 11/1.1)]}{E(X \wedge 22) - E(X \wedge 11)} = \frac{1.1(17.25 - 10)}{18.1 - 10.95} = 1.115.$$

Exercise 64 This exercise asks for quantities on a per loss basis. The expected value is

$$E(X) - E(X \wedge 100) = 1000 - 1000(1 - e^{-100/1000}) = 904.84.$$

To obtain the second moment, we need

$$\begin{aligned} E[(X \wedge 100)^2] &= \int_0^{100} x^2 0.001e^{-0.001x}dx + (100)^2 e^{-100/1000} \\ &= -e^{-0.001x}(x^2 + 2000x + 2{,}000{,}000)\Big|_0^{100} + 9048.37 \\ &= 9357.68. \end{aligned}$$

The second moment is

$$\begin{aligned} &E(X^2) - E[(X \wedge 100)^2] - 200E(X) + 200E(X \wedge 100) \\ &= 2(1000)^2 - 9357.68 - 200(1000) + 200(95.16) \\ &= 1{,}809{,}674.32 \end{aligned}$$

for a variance of $1{,}809{,}674.32 - 904.84^2 = 990{,}938.89$.

Exercise 65 Under the old plan the expected cost is $500/1 = 500$. Under the new plan the expected claim cost is K. The bonus is

$$B = \begin{cases} 0.5(500 - X), & X < 500 \\ 0.5(500 - 500), & X \geq 500 \end{cases}$$

which is 250 less a benefit with a limit of 500 and a coinsurance of 0.5. Therefore,

$$\begin{aligned} E(B) &= 250 - 0.5E(X \wedge 500) \\ &= 250 - 0.5\frac{K}{1}\left[1 - \left(\frac{K}{K+500}\right)\right] \\ &= 250 - \frac{K}{2} + \frac{K^2}{2K+1000}. \end{aligned}$$

The equation to solve is

$$500 = K + 250 - \frac{K}{2} + \frac{K^2}{2K+1000}$$

and the solution is $K = 354$.

Exercise 66 For year a, expected losses per claim are 2000 and thus 5000 claims are expected. Per loss, the reinsurer's expected cost is

$$\begin{aligned} E(X) - E(X \wedge 3000) &= 2000 - 2000\left[1 - \left(\frac{2000}{2000+3000}\right)\right] \\ &= 800 \end{aligned}$$

and therefore the total reinsurance premium is $1.1(5000)(800) = 4{,}400{,}000$. For year b, there are still 5000 claims expected. Per loss, the reinsurer's expected cost is

$$\begin{aligned} 1.05[E(X) - E(X \wedge 3000/1.05)] &= 1.05\left\{2000 - 2000\left[1 - \left(\frac{2000}{2000+3000/1.05}\right)\right]\right\} \\ &= 864.706 \end{aligned}$$

and the total reinsurance premium is $1.1(5000)(864.706) = 4{,}755{,}882$. The ratio is 1.0809.

Exercise 67 For this uniform distribution,

$$\begin{aligned} E(X \wedge u) &= \int_0^u x(0.00002)dx + \int_u^{50{,}000} u(0.00002)dx \\ &= 0.00001u^2 + 0.00002u(50{,}000 - u) \\ &= u - 0.00001u^2. \end{aligned}$$

From Theorem 5.13, the expected payment per payment is

$$\frac{E(X \wedge 25{,}000) - E(X \wedge 5{,}000)}{1 - F(5000)} = \frac{18{,}750 - 4{,}750}{1 - \frac{5{,}000}{50{,}000}} = 15{,}556.$$

Exercise 68 This is a combination of franchise deductible and policy limit, so none of the results apply. From the definition of this policy, the expected cost per loss is

$$\begin{aligned}
&\int_{50{,}000}^{100{,}000} xf(x)dx + 100{,}000[1 - F(100{,}000)] \\
&= \int_{0}^{100{,}000} xf(x)dx + 100,000[1 - F(100,000)] - \int_{0}^{50{,}000} xf(x)dx \\
&= E(X \wedge 100{,}000) - E(X \wedge 50{,}000) + 50{,}000[1 - F(50{,}000)].
\end{aligned}$$

Alternatively, it could be argued that this policy has two components. The first is an ordinary deductible of 50,000 and the second is a bonus payment of 50,000 whenever there is a payment. The first two terms above reflect the cost of the ordinary deductible and the third term is the extra cost of the bonus. Using the lognormal formulas in *LMA*, the answer is

$$\begin{aligned}
&e^{10.5}\Phi\left(\frac{\ln 100{,}000 - 10 - 1}{1}\right) + 100{,}000\left[1 - \Phi\left(\frac{\ln 100{,}000 - 10}{1}\right)\right] \\
&-e^{10.5}\Phi\left(\frac{\ln 50{,}000 - 10 - 1}{1}\right) \\
&= e^{10.5}\Phi(0.513) + 100{,}000[1 - \Phi(1.513)] - e^{10.5}\Phi(-0.180) \\
&= e^{10.5}(0.6959) + 100{,}000(0.0652) - e^{10.5}(0.4285) = 16{,}231
\end{aligned}$$

Exercise 69 We have

$$\begin{aligned}
60 - 0.6x &= \frac{\int_x^{100}(t - x)f(t)dt}{S(x)} \\
S(x)(60 - 0.6x) &= \int_x^{100}(t - x)f(t)dt \\
-f(x)(60 - 0.6x) - 0.6S(x) &= \int_x^{100} -f(t)dt = -S(x) \\
f(x)(60 - 0.6x) &= 0.4S(x) \\
\frac{f(x)}{S(x)} &= \frac{0.4}{60 - 0.6x}.
\end{aligned}$$

The third line follows from differentiating both sides with respect to x. The last line yields the hazard rate. Then,

$$\begin{aligned}
S(x) &= e^{-\int_0^x h(t)dt} \\
&= e^{-4\int_0^x (600-6t)^{-1}dt} \\
&= e^{(4/6)\ln(600-6t)|_0^x} \\
&= e^{(4/6)[\ln(600-6x)-\ln(600)]} \\
&= (1 - 0.01x)^{2/3}
\end{aligned}$$

and

$$f(x) = -S'(x) = \frac{2}{300}(1 - 0.01x)^{-1/3}.$$

From *LMA*, this can be recognized as a beta distribution with parameters $a = 1, b = 2/3$, and $\theta = 100$.

For the first problem, $S(65) = (0.35)^{2/3} = .4966$ and

$$\begin{aligned} E(Y) &= 1000 \sum_{j=0}^{34} 1.06^{-j}(0.35 - 0.01j)^{2/3}/0.4966 \\ &= 11{,}540.56. \end{aligned}$$

For the second problem,

$$\begin{aligned} E(Z) &= 1000 \int_0^{80} 1.06^{-x} f(20 + x) dx / S(20) \\ &= 1000 \int_0^{80} 1.06^{-x} \frac{2}{300}(0.8 - 0.01x)^{-1/3} dx / 0.86177 \\ &= 157.14 \end{aligned}$$

where numerical integration was used to get the answer.