`

Construction of Actuarial Models

An Introductory Guide for Actuaries and other Business Professionals

Fourth Edition

Michael Gauger, Ph.D.
Associate of the Society of Actuaries

Michael Hosking, MA
Associate of the Institute of Actuaries

BPP Professional Education
Farmington, CT

PROFESSIONAL EDUCATION™

Requests for permission should be addressed to:

BPP Professional Education
P.O. Box 283
Farmington, CT 06034

Manufactured in the United States of America

10 9 8 7 6 5 4 3 2 1

ISBN: 978-0-9816081-0-5

Preface to the Fourth Edition

Welcome to the fourth edition of this introductory guide to the Construction of Actuarial Models.

Based on our experience as professional educators, our aim when writing this text has been to produce a clear, practical and student-friendly guide in which theoretical derivations have been balanced with a helpful, structured approach to the material. We have supplemented the explanations with over 430 worked examples and practice questions to give students ample opportunity to see how the theory is applied. The result—we hope—is a thorough but accessible introduction to the construction and evaluation of actuarial models.

This text is of particular relevance to actuarial students who are preparing for Exam C of the Society of Actuaries, and Exam 4 of the Casualty Actuarial Society. Where possible, examples are set in an insurance or risk management context. For more information about an actuarial career, visit www.beanactuary.org or www.soa.org. Aspiring actuaries in the UK should visit www.actuaries.org.uk.

The numerical solutions to all of the end-of-chapter practice questions can be found at the end of the book. Detailed worked solutions to these practice questions can be downloaded free of charge from the BPP Professional Education website at **www.bpptraining.com.** Other useful study resources can also be found there.

For students preparing for the SOA's Exam C or the CAS's Exam 4, it is critically important to work as many exam-style questions as possible. For such preparation, this text should be used in conjunction with our supplemental Q&A Bank which contains hundreds of multiple choice questions (including relevant questions from past exams). Practice questions included in this book at the end of each chapter are designed to emphasize first principles and basic calculation whereas exam-style questions, such as those contained in the Q&A Bank, can be quite obtuse.

This edition could not have been completed without the helpful contributions of Julie Wilson and David Wilmot. Any errors in this text are solely our own.

We hope that you find this text helpful in your studies, wherever these may lead you.

Michael Hosking
November 2009

Table of contents

Introduction

Before we start the main subject matter in this text, we should take care of a little housekeeping.

Assumed knowledge

We assume that the reader has knowledge of calculus, probability theory, statistics, and the theory of survival models. The necessary background information from probability theory and actuarial models can be found in BPP's textbooks; *Probability* by Carr & Gauger and *Actuarial Models* by Gauger. You can visit our website and download a free sample chapter from each of these texts.

Notation and rounding

We have tried hard to ensure that all new notation is explained clearly. We sometimes use $\exp(x)$ in place of e^x, especially when this avoids complicated superscripts that might otherwise be difficult to read.

Rounding poses a particular dilemma. Our standard policy in this text has been to keep full accuracy within intermediate calculations even though an intermediate result may be shown as a rounded value. So, you may occasionally disagree with the last significant figure or two in a calculation if you calculate the result using the rounded values shown.

Solutions to practice questions

Short numerical answers to all of the end-of-chapter practice questions can be found at the end of this book. Detailed worked solutions to these practice questions can be downloaded free of charge from the BPP Professional Education website at **www.bpptraining.com** (look for "Text Question Solutions" on the left side menu). Other useful study resources can also be found there.

Errors in this text

If you find an error in this text, we'll be pleased to hear from you so that we can publish an errata list for students on our website and correct these errors in the next edition. Please email details of any errors to **examCsupport@bpptraining.com**. A current errata list is maintained in the Student Mailbag on the Exam C page of the BPP website **www.bpptraining.com**. Thank you.

1

Loss models

Overview

Over the first four chapters we will study models of aggregate loss distributions. These distributions are used to model things such as:

- the annual *amount* of claims paid by an auto insurer for a single policy
- the annual *number* of claims paid by an insurer on a group medical plan
- the annual payments of a reinsurance company to an insurance company where only a *part* of the insurer's claim payments are reimbursed by the reinsurer.

By the end of this chapter, we'll be able to:

- calculate the expected value and variance of such aggregate losses
- apply the concepts of mixing, splicing, truncating and censoring distributions to form new ones.

1.1 Introduction to loss models

Interest theory plays a significant role in models of *long-term* lines of business that include life insurance and annuities. This is because the premiums received by the insurer must be invested for a number of years before all of the annuity payments or death benefits are paid out. Actuarial present values are computed at issue to determine the annual benefit premium. Actuarial present values are also computed for each in-force policy to measure future liability with respect to the policy. Reserves must be set up so that the insurer has a clear picture of future liabilities for each line of business. The rate of interest used has a profound effect on pricing and reserving via the computation of both random present values and actuarial present values. Lower rates of interest will generally result in higher premiums and reserves.

In Chapters 1–4 of this text, we will develop models for *short-term* lines of business such as auto insurance, homeowner's insurance, or health insurance. These policies (contracts) typically cover a one-year time period. When the policy comes up for renewal after the year is over, the insurer determines a new premium rate that is based on recent claim experience and perhaps some regulatory requirements. Since premium income is often held by the insurer for a relatively short period of time before claims occur, the interest it could earn can usually be ignored.

Levels of loss

The term **loss** is used at several distinct levels:

- *Insured*

  It could be used to refer to the actual out of pocket expense as a result of some insured random event such as an automobile accident. It could also be the cost of medical treatment and compensation for "pain" resulting from an attack by some insured's pit bull.

  In general it refers to a monetary measure of what is "lost" by the insured as a result of an insured event. This type of loss is often referred to as the **ground up loss**.

- *Insurer*

  From an insurer's point of view the money it pays out in **claims** is viewed as a loss against the premium income received from policyholders.

  A ground up loss of a policyholder might be completely covered by the insurance. However, it is more common that the ground up loss is subject to a **deductible** amount or a **limit** (to be covered in Chapter 2). In this case, part of the ground up loss (up to a maximum of the deductible) is retained by the policyholder, and any remainder of the ground up loss is paid by the insurer as a claim. In other words, there is a sharing of each ground up loss between the policyholder and the insurer.

  The loss amount is split into the sum of the amount retained by the insured and the amount reimbursed to the insured by the insurer.

- *Reinsurer*

  Now move up to the next level and consider the total annual claims paid by the insurer for this line of business (the insurer's share of all ground up losses). To limit the risk that aggregate claims might far exceed premium income from the line of business, the insurer might seek a reinsurance treaty (policy) whereby its claims are shared with a reinsurer in return for part of the insurer's premium income. Effectively, the insurer takes out its own insurance with the reinsurer.

Part of the claims are retained by the insurer and part of the claims are **ceded** to the reinsurer (*ie* become the legal responsibility of the reinsurer).

So the insurer might enter into a treaty where the insurer pays a reinsurance premium to the reinsurer, who is then obligated to pay part of the insurer's claims. In this way, these payments represent losses from the reinsurer's point of view.

In order to create models for such aggregate "losses" of insurers and reinsurers, and the loss sharing arrangements, we will need to use some relatively sophisticated probability results. Some of these results are not typically part of a first course in probability theory. Others are covered, but not in sufficient depth.

To remedy this situation we will introduce the necessary probability theory in this chapter at the same time as we discuss the individual risk model (IRM) and the collective risk model (CRM).

The theory in this chapter plays a pivotal role relative to Chapters 2-4. Understanding the loss model terminology and the probability results introduced here is absolutely essential for understanding Chapters 2-4. Take the time to read this current chapter several times. The reward for a thorough understanding of this material will be that Chapters 2-4 will then be a lot easier to work through.

1.2 Moment and probability generating functions

The moment generating function and probability generating function associated with a random variable X are real valued functions of a real variable. They are useful for several purposes:

- calculating moments of the distribution
- identifying the distribution of a sum of independent random variables from the same parametric family.

Let's begin with their definitions.

Moment generating function

The **moment generating function** of a random variable X, denoted by $M(t)$ or $M_X(t)$, is a real valued function of the real variable *t* defined as:

$$M_X(t)=E\left[e^{tX}\right]=\sum_x e^{tx}\Pr(X=x) \qquad \text{(discrete case)}$$

$$=\int_{-\infty}^{\infty} e^{tx} f_X(x)\,dx \qquad \text{(continuous case)}$$

The domain of this function is the set of all values of *t* such that the sum or integral exists. There is the possibility that for some *t* the sum might be a divergent infinite series or the integral might be a divergent improper integral. The moment generating function is useful only when it is defined on an open interval containing zero.

For example, if X is a discrete random variable with $\Pr(X=i)=1/3$ for $i=1,2,3$, then the moment generating function is:

$$M_X(t)=\sum_x e^{tx}\Pr(X=x) = \frac{1}{3}e^t+\frac{1}{3}e^{2t}+\frac{1}{3}e^{3t} \quad \text{for} \quad -\infty < t < \infty$$

Example 1.1

Suppose that the discrete random variable N has a probability function given by:

$$\Pr(N=k) = e^{-\lambda}\,\frac{\lambda^k}{k!} \quad \text{where } k=0,1,2,\ldots \text{ and } \lambda>0$$

(This is the **Poisson distribution** with parameter λ which will be covered extensively in Chapter 3.)

Determine the moment generating function.

Solution

It takes no effort to copy down the definition:

$$M_N(t)=E\left[e^{tN}\right]=\sum_{k=0}^{\infty} e^{tk}\Pr(N=k)=\sum_{k=0}^{\infty} e^{tk}\,e^{-\lambda}\,\frac{\lambda^k}{k!}$$

However, in this form the moment generating function is virtually useless. To make a generating function truly useful you must be able to write the summation in a more compact form. The result we need here is the **Taylor series for the exponential function**:

$$e^x=\sum_{k=0}^{\infty}\frac{x^k}{k!}=1+x+\frac{x^2}{2!}+\frac{x^3}{3!}+\cdots \qquad (\text{converges for } -\infty<x<\infty)$$

Applying this result to our generating function definition we have:

$$M_N(t)=\sum_{k=0}^{\infty} e^{tk}\,e^{-\lambda}\,\frac{\lambda^k}{k!}=e^{-\lambda}\sum_{k=0}^{\infty}\frac{\left(\lambda e^t\right)^k}{k!}=e^{-\lambda}\,e^{\lambda e^t}=e^{\lambda\left(e^t-1\right)} \quad \text{for } -\infty<t<\infty$$

♦♦

Example 1.2

Suppose that the continuous random variable X has a probability density function given by:

$$f_X(x)=\frac{e^{-x/\theta}}{\theta} \quad \text{for } x>0 \text{ and } \theta>0$$

(This is the **exponential distribution** with parameter θ that will be covered extensively in Chapter 2.)

Determine the moment generating function.

Solution

From the definition, we have:

$$M_X(t)=\int_{-\infty}^{\infty} e^{tx} f_X(x)\,dx=\int_0^{\infty} e^{tx}\,\frac{e^{-x/\theta}}{\theta}\,dx$$

Once again, to make the function useful we need to evaluate the integral so as to obtain a closed form expression in the real variable t:

$$M_X(t)=\int_0^{\infty} e^{tx}\,\frac{e^{-x/\theta}}{\theta}\,dx=\frac{1}{\theta}\int_0^{\infty} e^{-x\left(\theta^{-1}-t\right)}\,dx$$

In general, we have:

$$\int_0^\infty e^{-ax}\,dx = \lim_{b\to\infty}\int_0^b e^{-ax}\,dx = \lim_{b\to\infty}\left(\frac{1-e^{-ab}}{a}\right) = \frac{1-0}{a} = \frac{1}{a} \quad \text{if } a>0$$

The integral diverges if $a \le 0$. Comparing this result to the last form of the generating function, we see that:

$$M_X(t) = \frac{1}{\theta}\int_0^\infty e^{-x\left(\theta^{-1}-t\right)}\,dx = \frac{1}{\theta}\times\frac{1}{\theta^{-1}-t} = \left(1-\theta t\right)^{-1} \quad \text{if } t<\theta^{-1}$$

So the generating function has a simple closed form. In contrast with the result in Example 1.1, the generating function in this case is not defined for all real numbers. ♦♦

The fact that we saw a smaller domain for the moment generating function in Example 1.2 than we found for the distribution in Example 1.1 is no real drawback. The key thing is that the function is defined in an interval that includes zero. The following is a summary of key properties of the moment generating function. We won't include the proofs here but they are readily available from probability texts.

Properties of moment generating functions

1. $E\left[X^k\right] = M_X^{(k)}(0)$ provided the k-th derivative exists at zero.

2. If $M_X(t) = M_Y(t)$, then X and Y are identically distributed.

3. If $X_1, X_2, \ldots, X_n$ are independent random variables, then we have:

$$M_{\sum a_i X_i}(t) = M_{X_1}(a_1 t)\cdots M_{X_n}(a_n t)$$

Example 1.3

Compute the mean and variance for the exponential distribution in Example 1.2.

Solution

Using the result of Example 1.2 and Property 1 we have:

$$M_X(t) = (1-\theta t)^{-1} \;\Rightarrow\; M_X'(t) = \theta(1-\theta t)^{-2} \;\Rightarrow\; M_X''(t) = 2\theta^2(1-\theta t)^{-3}$$

$$E[X] = M_X'(0) = \theta \quad , \quad E\left[X^2\right] = M_X''(0) = 2\theta^2$$

$$\Rightarrow\; \operatorname{var}(X) = E[X^2] - (E[X])^2 = 2\theta^2 - (\theta)^2 = \theta^2$$

♦♦

Example 1.4

Show that the sum of two independent Poisson distributed random variables follows a Poisson distribution whose parameter is the sum of the two component parameters.

Solution

Suppose that $M_{N_i} = e^{\lambda_i\left(e^t-1\right)}$ is the moment generating function for $i=1,2$. (See Example 1.1.)

Assuming that these variables are independent, Property 3 results in:

$$M_{N_1+N_2}(t) = M_{N_1}(t)M_{N_2}(t) = e^{\lambda_1\left(e^t-1\right)}\, e^{\lambda_2\left(e^t-1\right)} = e^{(\lambda_1+\lambda_2)\left(e^t-1\right)}$$

This function is exactly the moment generating function for a Poisson distribution with parameter $\lambda_1+\lambda_2$. So by Property 2, it follows that N_1+N_2 follows a Poisson distribution with parameter $\lambda_1+\lambda_2$. ♦♦

Probability generating function

The **probability generating function** associated with a discrete random variable *X* is defined by:

$$P_X(t) = E\left[t^X\right] \text{ for all } t>0 \text{ such that the expected value exists}$$

It is closely related to the moment generating function:

$$P_X(t) = E\left[t^X\right] = E\left[e^{\ln(t)X}\right] = M_X\left(\ln(t)\right)$$

Equivalently, we have:

$$P_X\left(e^t\right) = M_X(t)$$

The probability generating function has properties similar to those of the moment generating function. Once again, proofs are beyond the scope of this course.

Properties of probability generating functions

1. $P_X^{(k)}(1) = E\left[X(X-1)\cdots(X-k+1)\right]$ if the derivative exists.

2. If $P_X(t) = P_Y(t)$, then *X* and *Y* are identically distributed.

3. If $X_1, X_2, \ldots, X_n$ are independent random variables, then we have:

$$P_{\sum X_i}(t) = P_{X_1}(t)\cdots P_{X_n}(t)$$

4. $P_X(t)=M_X\left(\ln(t)\right)$

Our main use of the probability generating function will occur in Chapter 3.

Example 1.5

Compute the probability generating function for the Poisson distribution in Example 1.1.

Solution

$$M_X(t) = e^{\lambda\left(e^t-1\right)} \Rightarrow P_X(t) = M_X\left(\ln(t)\right) = e^{\lambda\left(e^{\ln(t)}-1\right)} = e^{\lambda(t-1)}$$ ♦♦

1.3 Sums of independent random variables and the IRM

Sums of independent random variables will occur in several settings over the course of Chapters 1-4. For example:

- If a portfolio (line of business) consists of n independent risks (policies), then the annual claim frequency for the portfolio is the sum of the annual claim frequencies for the individual policies.
- The aggregate annual claim amount is usually modeled as a sum of independent individual claim amounts.

The individual risk model (IRM)

Suppose that a portfolio consists of n risks (policies) whose total annual losses $X_1, X_2, \ldots, X_n$ are independent and identically distributed like X. The aggregate annual loss for the entire portfolio is represented by the sum

$$S = X_1 + X_2 + \cdots + X_n$$

This method of modeling the aggregate annual loss is known as the **Individual Risk Model** since it focuses on the losses from the individual policies (risks).

The exact distribution of a sum of independent random variables

A very important basic problem is therefore to compute the distribution of a sum of independent random variables from knowledge of the distributions of the component independent random variables. There are three main options as to how this might be done:

1. If the random variables are from the same parametric family of distributions, then the distribution of their sum will often be from this family as well. Moment generating functions are very useful in this regard as we saw in Example 1.4.
2. Recursively using the method of convolutions (see below).
3. Approximately using the Central Limit Theorem (see below).

The Method of Convolutions

If $f_X(x)$ and $f_Y(y)$ are the probability functions or PDF's for independent random variables X and Y, then the probability function or PDF for the sum $S = X + Y$ is referred to as the **convolution of** f_X **with** f_Y and is sometimes denoted by:

$$f_X * f_Y(s)$$

It is an elementary argument to establish the following relations:

$$f_X * f_Y(s) = \Pr(X + Y = s) = \sum_x \Pr(X = x\ , Y = s - x) \qquad \text{(discrete case)}$$

$$= \sum_x \Pr(X = x)\Pr(Y = s - x) \qquad \text{(independence)}$$

$$= \sum_x f_X(x) f_Y(s - x)$$

$$f_X * f_Y(s) = \int_{-\infty}^{\infty} f_X(x) f_Y(s - x)\, dx \qquad \text{(continuous case)}$$

In both loss theory and survival models our random variables can *only take on non-negative values*. In this case there is a small adjustment to the formulas above since both x and $s-x$ must be non-negative:

Convolution formulas

$$f_X * f_Y(s) = \sum_{0 \le x \le s} f_X(x) f_Y(s-x) \qquad \text{(discrete non-negative random variables)}$$

$$f_X * f_Y(s) = \int_0^s f_X(x) f_Y(s-x)\,dx \qquad \text{(continuous non-negative random variables)}$$

The summation formula above should be pretty intuitive. The possible ways of obtaining $X+Y=s$ consist of all combinations where x is between 0 and s, and where y is equal to $s-x$.

Theoretically this technique could be used recursively to compute the distribution of a sum of more than 2 independent variables. For example, if $X_1, X_2, \ldots, X_n$ are independent and identically distributed like X, then the PDF of the sum $S = X_1 + \cdots + X_n$, denoted by $f_X^{*n}(s)$, is computed recursively by the rule:

$$\begin{aligned} f_X^{*k+1}(s) &= f_X^{*k} * f_X(s) \qquad k \ge 1 \\ &= \sum_{0 \le x \le s} f_X^{*k}(x) f_X(s-x) \end{aligned}$$

where $f_X^{*1}(x) = f_X(x)$.

By hand these calculations quickly become cumbersome. They are not as difficult using a computer but, if n is large, even computer calculations can become impractical.

Example 1.6

Assume that the discrete random variables X_1, X_2, X_3 are independent and identically distributed like X where $f_X(1) = 0.7$, $f_X(2) = 0.3$.

Compute f_X^{*2} and f_X^{*3}.

Solution

The possible values of $X_1 + X_2$ are 2, 3, and 4:

$$f_X^{*2}(2) = \Pr(X_1 = 1)\Pr(X_2 = 1) = 0.49$$

$$f_X^{*2}(3) = \Pr(X_1 = 1)\Pr(X_2 = 2) + \Pr(X_1 = 2)\Pr(X_2 = 1) = 2 \times 0.7 \times 0.3 = 0.42$$

$$f_X^{*2}(4) = \Pr(X_1 = 2)\Pr(X_2 = 2) = 0.09$$

The possible values of $X_1+X_2+X_3$ are 3, 4, 5, and 6:

$$f_X^{*3}(3) = \Pr(X_1+X_2=2)\Pr(X_3=1) = 0.49\times 0.7 = 0.343$$

$$f_X^{*3}(4) = \Pr(X_1+X_2=2)\Pr(X_3=2) + \Pr(X_1+X_2=3)\Pr(X_3=1)$$
$$= 0.49\times 0.3 + 0.42\times 0.7 = 0.441$$

$$f_X^{*3}(5) = \Pr(X_1+X_2=3)\Pr(X_3=2) + \Pr(X_1+X_2=4)\Pr(X_3=1)$$
$$= 0.42\times 0.3 + 0.09\times 0.7 = 0.189$$

$$f_X^{*3}(6) = \Pr(X_1+X_2=4)\Pr(X_3=2) = 0.09\times 0.3 = 0.027$$ ♦♦

Example 1.7

Assume X_1, X_2 are independent and identically distributed like X and that $f_X(x)=1$ for $0<x<1$.

Compute f_X^{*2}.

Solution

The possible values of the sum lie between 0 and 2. For s between 0 and 1 we have:

$$f_X^{*2}(s) = \int_0^s \underbrace{f_X(x)\,f_Y(s-x)}_{\substack{\text{both factors are 1}\\ \text{since } 0<x,\, s-x<1}}dx = s \quad , \quad 0<s\le 1$$

For s between 1 and 2, the integrand $f_X(x)\,f_Y(s-x)$ is non-zero only when $s-1<x<1$. So we have:

$$f_X^{*2}(s) = \int_{s-1}^1 \underbrace{f_X(x)\,f_Y(s-x)}_{\substack{\text{both factors are 1}\\ \text{since } 0<x,\, s-x<1}}dx = 2-s \quad , \quad 1\le s<2$$ ♦♦

Applying the Central Limit Theorem

The number of policies n will typically be large, so that computing the exact distribution of S by the recursive method of convolutions will be impossible by hand, and perhaps even impractical using a computer. In this case, we typically apply the **Central Limit Theorem** and approximate the distribution of S by a normal distribution with mean $\mu = nE[X]$ and variance $\sigma^2 = n\,\mathrm{var}(X)$.

Example 1.8

Suppose that the annual loss for an individual policy has the following distribution:

$$\Pr(X=0)=0.75\ ,\ \Pr(X=1)=0.15\ ,\ \Pr(X=2)=0.10$$

Suppose that there are $n=100$ independent policies in a portfolio whose annual losses are each distributed like X.

Determine the expected annual loss for the portfolio, the variance in annual loss for the portfolio, and the approximate 90th percentile of aggregate annual loss.

Solution

We start by calculating the mean and variance of X:

$$E[X] = \sum_{x=0}^{2} x \Pr(X = x) = 0.15 + 0.20 = 0.35$$

$$E\left[X^2\right] = \sum_{x=0}^{2} x^2 \Pr(X = x) = 0.15 + 0.40 = 0.55$$

$$\Rightarrow \quad \text{var}(X) = 0.55 - 0.35^2 = 0.4275$$

The aggregate annual loss $S = X_1 + \cdots + X_{100}$ is approximately normally distributed with:

$$E[S] = 100\,E[X] = 35 \quad , \quad \text{var}(S) = 100\,\text{var}(X) = 42.75$$

The 90th percentile of the standard normal distribution is 1.282. Since S is approximately normal in distribution, the approximate 90th percentile of S is:

$$E[S] + 1.282\sqrt{\text{var}(S)} = 35 + 1.282\sqrt{42.75} = 43.38$$ ♦♦

The individual risk model (IRM)

Aggregate annual claims are $S = X_1 + X_2 + \cdots + X_n$ where $X_1, \ldots, X_n$ are independent and identically distributed like X, where X is the model for total annual losses against a single policy (risk).

1. $E[S] = n\,E[X] \quad , \quad \text{var}(S) = n\,\text{var}(X)$

2. If $n \geq 50$, then:

$$\Pr(S \leq F) = \Pr\left(\frac{S - E[S]}{\sqrt{\text{var}(S)}} \leq \frac{F - E[S]}{\sqrt{\text{var}(S)}}\right) \approx \Phi\left(\frac{F - E[S]}{\sqrt{\text{var}(S)}}\right) = \Phi\left(\frac{F - nE[X]}{\sqrt{n\,\text{var}(X)}}\right)$$

where $\Phi(x)$ is the cumulative distribution function of the standard normal distribution.

3. The approximate $100(1-\alpha)$ th percentile of S is:

$$E[S] + z_\alpha\sqrt{\text{var}(S)} = n\,E[X] + z_\alpha\sqrt{n\,\text{var}(X)}$$

where $\alpha = \Pr\left(N(0,1) > z_\alpha\right) = 1 - \Phi(z_\alpha)$.

There are several *drawbacks* to the IRM:

- If S is assumed to be approximately normal in distribution, a small (non-zero) amount of probability will be assigned to the impossible event $S < 0$.
- In practice, the graph of the PDF of most aggregate loss distributions is **skewed to the right** (there is a significant probability attached to large losses), meaning that it has a long thin tail of area at the right end. If the distribution of annual losses is skewed to the right and the central limit theorem is used to fit a normal PDF (which isn't skewed) then right tail probabilities such as $\Pr(S > F)$ can be badly underestimated by the approximation:

$$\Pr(S > F) \approx 1 - \Phi\left(\frac{F - nE[X]}{\sqrt{n\,\text{var}(X)}}\right)$$

1.4 The Double Expectation Theorem (DET)

The double expectation theorem is a device that is employed when conditional moments of a random variable are easier to compute than the unconditional moments. Once we've stated it we will show how it can be applied. Its proof appears in the appendix to Chapter 1.

Theorem 1.1: The Double Expectation Theorem (DET)

For any random variables X and Y where X may depend on Y, we have:

$$E[X] = E\big[E[X|Y]\big]$$
$$\text{var}(X) = E\big[\text{var}(X|Y)\big] + \text{var}\big(E[X \mid Y]\big)$$

The abstract nature of these relations obscures their meaning. The following example from the theory of contingent payment models should help you understand how to use it.

Example 1.9

Suppose that females in a certain population have a constant force of mortality equal to $\mu_F = 0.015$, and that males in this population have a constant force of mortality equal to $\mu_M = 0.020$.

Let $Y = \bar{a}_{\overline{T(x)}|}$ be the random present value variable for a continuous life annuity of 1 per year for a randomly selected member of this population. Assuming that the force of interest is $\delta = 0.05$, and that 55% of the population are female, determine the expected value and variance of Y.

Solution

We can use some constant force formulas to help us here.

If $Y = \bar{a}_{\overline{T(x)}|}$ and the life (x) is a *male*, we have:

$$E[Y] = \frac{1}{\mu_M + \delta} = \frac{1}{0.07} = 14.28571$$
$$\text{var}(Y) = \frac{1}{\delta^2}\left({}^2\bar{A}_x - \left(\bar{A}_x\right)^2\right) = \frac{1}{\delta^2}\left(\frac{\mu_M}{\mu_M + 2\delta} - \left(\frac{\mu_M}{\mu_M + \delta}\right)^2\right) = 34.01361$$

If the life (x) is a *female*, then the corresponding results are:

$$E[Y] = \frac{1}{\mu_F + \delta} = \frac{1}{0.065} = 15.38462$$
$$\text{var}(Y) = \frac{1}{\delta^2}\left({}^2\bar{A}_x - \left(\bar{A}_x\right)^2\right) = \frac{1}{\delta^2}\left(\frac{\mu_F}{\mu_F + 2\delta} - \left(\frac{\mu_F}{\mu_F + \delta}\right)^2\right) = 30.87214$$

In other words, if we are *given* the gender of the randomly selected life, then we can compute the expected value and variance of the random present value variable. So what we need to do first is create an indicator of gender.

Let the **indicator variable** I be defined by:

$$I = \begin{cases} 1 & \text{for a male life} \\ 2 & \text{for a female life} \end{cases}$$

where

$$\Pr(I=1) = 0.45 \ , \ \Pr(I=2) = 0.55$$

The expected value and variance calculations can now be viewed as conditional moments:

I	Probability	$E[Y\|I]$	$\text{var}(Y\|I)$
1	0.45	14.28571	34.01361
2	0.55	15.38462	30.87214

Now apply the first part of the DET:

$$E[Y]=E\big[E[Y|I]\big] = \underbrace{0.45 \times 14.28571 + 0.55 \times 15.38462}_{\text{a weighted average of the expected value for the two "subgroups"}} = 14.89011$$

Apply the second part of the DET to obtain the variance:

$$\text{var}(Y) = \underbrace{\text{var}\big(E[Y|I]\big)}_{\substack{\text{variability between}\\ \text{subgroup means}}} + \underbrace{E\big[\text{var}(Y|I)\big]}_{\substack{\text{average variability}\\ \text{within the subgroups}}}$$

$$=\left(0.45 \times 14.28571^2 + 0.55 \times 15.38462^2 - 14.89011^2\right)$$
$$+ 0.45\times 34.01361 + 0.55\times 30.87214$$

$$= \underbrace{0.29888}_{\substack{\text{"between"}\\ \text{subgroups}}} + \underbrace{32.28580}_{\substack{\text{"within"}\\ \text{subgroups}}} = \underbrace{32.58468}_{\text{total}}$$ ♦♦

We will return to this example later and look at how to tackle it in a different way. For now, just notice that intuitively, the PDF for the future lifetime of a randomly selected life should be a weighted average of the PDF for the exponential future lifetime of a male and the exponential future lifetime of a female:

$$f_T(t) = 0.45 f_M(t) + 0.55 f_F(t) = 0.45\times 0.02e^{-0.02t} + 0.55\times 0.015e^{-0.015t} \quad , 0 < t < \infty$$

The moments of Y would then be:

$$E\left[Y^k\right] = E\left[\left(\bar{a}_{\overline{T}|}\right)^k\right] = \int_0^\infty \left(\frac{1-v^t}{\delta}\right)^k f_T(t)\,dt$$

1.5 Compound sums and the Collective Risk Model (CRM)

In the Individual Risk Model for aggregate annual losses, the focus was on the individual policies that make up the portfolio. We let X be the model for the **total annual loss** from a single policy. If there are n independent policies in the portfolio, each with total annual loss distributed like X, then the aggregate annual loss for the portfolio was modeled by:

$$S = X_1 + \cdots + X_n$$

In the **Collective Risk Model** we ignore the individual policies that generate the losses. We assume that over the course of a year there are a random number of losses, N, that are generated by a portfolio of risks. The **individual loss amounts** $Y_1, \ldots, Y_N$ are assumed to be independent and identically distributed like Y. We assume that $Y > 0$. Furthermore, the loss amounts are assumed to be independent of N. Under the Collective Risk Model, the aggregate annual loss for the portfolio is then modeled by the **compound sum** or **random sum:**

$$S = Y_1 + \cdots + Y_N$$

The random variable N is known as the **frequency** component of the compound sum. It is a **counting distribution** that has 0, 1, 2, and so on, as its possible values. The individual loss amount model Y is referred to as the **severity** component. The number of risks in the portfolio is hidden from view, but it will clearly affect the distribution of N. A greater number of policies will be reflected in a greater annual frequency of losses.

Here is the connection between these two modeling methods. The ith policy experiences a random number N_i of *individual* loss amounts where $N_i \geq 0$. The frequency of annual losses for the entire portfolio, N, is the sum of the annual loss frequencies over the various policies:

$$N = N_1 + \cdots + N_n$$

Each individual policy's total annual loss is therefore the sum of a random number N_i of individual loss amounts each distributed like Y. So we have:

$$X_i \sim Y_1 + \cdots + Y_{N_i}$$

In other words, X_i is itself a compound sum. It can happen that $X_i = 0$ when $N_i = 0$.

Thus the difference between the IRM and the CRM is how the individual loss amounts are grouped:

- In the CRM the aggregate annual loss S is computed as the sum of the individual loss amounts Y_i as they occur.
- For the IRM, these individual losses Y_i are first grouped into total annual losses from the n individual policies to determine $X_1, \ldots, X_n$. Then the various X_j are summed to determine S.

A key advantage of the CRM is that we can write an exact formula for f_S in terms of f_N for the frequency component and f_Y for the severity component. We can adjust these two individual models independently. For example:

- Increased exposure for the insurer (more policies in the portfolio or more coverages being added to the individual policies) can be handled by adjusting parameters of the frequency model N. (This is considered in Chapter 3.)
- One can begin by assuming that Y is an individual ground up loss experienced by a policyholder. The insurer might wish to apply a deductible amount, a limit, or a coinsurance (explained later) factor to the ground up loss in order to limit the severity of claim payments (*ie* losses). Or the insurer might want to model the effect of loss-inflation on S from one year to the next. These tasks can be accomplished by adjusting the distribution of the severity component Y. (This is considered in Chapter 2.)

In this section we want to derive the basic relations between the distributions of S, N, and Y for the CRM (or random sum model) for aggregate annual losses of an insurer or reinsurer. More advanced problems will be deferred until we have studied frequency and severity models in greater detail.

The difficulty with analyzing the random sum $S=Y_1+\cdots+Y_N$ is the random number of terms being summed. But if we knew that $N=k$ (*ie* if we were *given* $N=k$), then the sum is easily dealt with:

$$E\left[S|N=k\right]=E\left[Y_1+\cdots+Y_k\right]=k\,E[Y]$$
$$\text{var}\left(S|N=k\right)=\text{var}\left(Y_1+\cdots+Y_k\right)=k\,\text{var}(Y)$$
$$f_S\left(s \mid N=k\right)=f_Y^{*k}(s)$$

These relations are true for all k. The first two are typically rewritten in the following form that emphasizes that the conditional mean and variance of S are *linear functions* of N:

$$E\left[S \mid N\right]=E[Y]\,N$$
$$\text{var}\left(S \mid N\right)=\text{var}(Y)\,N$$

As you look at these formulas keep in mind that $E[Y]$ and $\text{var}(Y)$ are fixed real numbers and that N is a random variable.

Now you can begin to appreciate the beneficial nature of the double expectation theorem. In view of our conditional mean and variance formulas above, it is the ideal tool for relating moments of S to moments of the frequency and severity distributions.

Theorem 1.2

Suppose that $S=Y_1+\cdots+Y_N$ where the various Y_i are independent and identically distributed like Y where Y is non-negative. Suppose also that the individual loss amounts Y_i are independent of the annual loss frequency N.

Then it follows that:

(i) $E[S]=E[N]\,E[Y]$

(ii) $\text{var}(S)=E[N]\,\text{var}(Y)+\left(E[Y]\right)^2\,\text{var}(N)$

(iii) $f_S(s)=\sum_{k=0}^{\infty}\Pr(N=k)\,f_Y^{*k}(s)$

(iv) $M_S(t)=M_N\left(\ln\left(M_Y(t)\right)\right)$ and $P_S(t)=P_N\left(P_Y(t)\right)$

(v) If Y is a continuous random variable, then $\Pr(S=0)=\Pr(N=0)$. But if Y is discrete, we have $\Pr(S=0)=P_N\left(\Pr(Y=0)\right)$ if $\Pr(Y=0)>0$, or $\Pr(S=0)=\Pr(N=0)$ otherwise.

Proof

(i) We have seen in the preceding discussion that:

$$E\left[S \mid N\right]=E[Y]\,N$$
$$\text{var}\left(S \mid N\right)=\text{var}(Y)\,N$$

Now apply the double expectation theorem:

$$E[S]=E\left[E\left[S \mid N\right]\right]=E\left[E[Y]N\right]=E[N]\,E[Y]$$

(ii) Similarly, we have:

$$\begin{aligned}\operatorname{var}(S) &= E\left[\operatorname{var}(S \mid N)\right] + \operatorname{var}\left(E\left[S \mid N\right]\right)\\ &= E\left[\operatorname{var}(Y)N\right] + \operatorname{var}\left(E[Y]\,N\right)\\ &= E[N]\operatorname{var}(Y) + \left(E[Y]\right)^2 \operatorname{var}(N)\end{aligned}$$

(iii) Let's just consider the case when Y is a discrete random variable. So we have $f_Y(y) = \Pr(Y = y)$. Since S is a sum of Y's, it is also a discrete random variable. Therefore, we also have $f_S(s) = \Pr(S = s)$.

There are many smaller events making up the event $\{S = s\}$:

$$\{S = s\} = \bigcup_{k=0}^{\infty} \underbrace{\{N = k \text{ and } Y_1 + Y_2 + \cdots + Y_k = s\}}_{\text{there are } k \text{ terms adding up to } s}$$

The events on the right side are mutually exclusive. Since N and the various Y_i are independent, we have:

$$\begin{aligned}f_S(s) = \Pr(S = s) &= \Pr\left(\bigcup_{k=0}^{\infty}\{N = k \text{ and } Y_1 + Y_2 + \cdots + Y_k = s\}\right)\\ &= \sum_{k=0}^{\infty} \Pr\left(N = k \text{ and } Y_1 + Y_2 + \cdots + Y_k = s\right)\\ &= \sum_{k=0}^{\infty} \Pr(N = k)\underbrace{\Pr\left(Y_1 + Y_2 + \cdots + Y_k = s\right)}_{f_Y^{*k}(s)}\end{aligned}$$

(iv) The formula $P_S(t) = P_N\left(P_Y(t)\right)$ can also be established with the help of the Double Expectation Theorem:

$$\begin{aligned}P_S(t) = E\left[t^S\right] = E\left[E\left[t^S \mid N\right]\right] &= E\left[E\left[t^{Y_1 + \cdots + Y_N} \mid N\right]\right]\\ &= E\left[E\left[t^{Y_1 + \cdots + Y_N}\right]\right] && \text{(the variables } Y_1, Y_2, \ldots, Y_N, N \text{ are independent)}\\ &= E\left[E\left[t^{Y_1}\right]E\left[t^{Y_2}\right]\cdots E\left[t^{Y_N}\right]\right] && \text{(the variables } Y_1, Y_2, \ldots, Y_N \text{ are independent)}\\ &= E\left[E\left[t^{Y}\right]E\left[t^{Y}\right]\cdots E\left[t^{Y}\right]\right] && \text{(each of } Y_1, Y_2, \ldots, Y_N \text{ is distributed like } Y)\\ &= E\left[P_Y(t)\cdots P_Y(t)\right] = E\left[\left(P_Y(t)\right)^N\right] = P_N\left(P_Y(t)\right)\end{aligned}$$

The formula $M_S(t) = M_N\left(\ln\left(M_Y(t)\right)\right)$ is now easily derived since $M_N(t) = P_N\left(e^t\right)$:

$$M_S(t) = P_S\left(e^t\right) = P_N\left(P_Y\left(e^t\right)\right) = P_N\left(M_Y(t)\right) = P_N\left(e^{\ln\left(M_Y(t)\right)}\right) = M_N\left(\ln\left(M_Y(t)\right)\right)$$

(v) Consider the formula derived in (iii). We can see that if Y is a *continuous* random variable then the terms on the right side with $k = 1, 2, \ldots$ form a weighted sum of continuous random variables, since a sum of continuous random variables is continuous. So for this part of the distribution of S there is no probability at zero. However, in the term with $k = 0$, the symbol $f_Y^{*0}(s)$ is the PDF of an empty sum of terms distributed like Y. In other word, we have:

$f_Y^{*0}(0) = 1$ (the PDF of a random variable that is certain to be zero)

So $\Pr(S=0)=\Pr(N=0)$.

In the case where Y is *discrete* it is also true that $Y_1+Y_2+\cdots+Y_k$ is discrete. As a result, it follows that $f_Y^{*k}(0)=\Pr(Y_1+Y_2+\cdots+Y_k=0)$. Since all of the Y_i are non-negative, it follows that:

$$\begin{aligned}\Pr(Y_1+Y_2+\cdots+Y_k=0)&=\Pr(\text{all } Y_i \text{ are zero})\\&=\Pr(Y_1=0)\Pr(Y_2=0)\cdots\Pr(Y_k=0) &&\text{(independence)}\\&=(\Pr(Y=0))^k &&(\text{identically distributed like } Y)\end{aligned}$$

Finally, we have:

$$\Pr(S=0)=f_S(0)=\sum_{k=0}^{\infty}\Pr(N=k)\Pr(Y=0)^k=\sum_{k=0}^{\infty}\Pr(N=k)t^k \quad \text{where } t=\Pr(Y=0)$$

$$=E\left[t^N\right]=P_N(t)=P_N(\Pr(Y=0)) \qquad \square$$

There are a number of important observations to make regarding this theorem listing the basic properties of random sums:

- Property (i) is a fairly intuitive idea. It says that the expected value of a random sum is equal to the product of the expected number of terms and the expected value of a term. Notice that if $\Pr(N=n)=1$ (in other words there are always $N=n$ terms where n is a fixed number), then property (i) is exactly like the well-known formula $E[Y_1+\cdots+Y_n]=nE[Y]$.

- Property (ii) should also seem reasonable. In a random sum there are two sources of variability – variance in the number of terms and variance in the amounts of the individual terms.

  The term $E[N]\text{var}(Y)$ is proportional to the variance in term amounts and reflects the second source. The term $(E[Y])^2\text{var}(N)$ is proportional to variance in the number of terms.

  Once again, if $\Pr(N=n)=1$, then the compound sum variance formula reduces to the familiar result: $\text{var}(Y_1+\cdots+Y_n)=n\,\text{var}(Y)$.

- Property (iii) shows how to combine the functions f_N and f_Y to obtain an exact formula for f_S. However, it is difficult to use in practice since it is usually not easy to compute the convolution f_Y^{*k}.

  Nevertheless, there is a simple combinatorial idea underlying property (iii) that can be used to compute $\Pr(S=0)$, $\Pr(S=1)$, … , and so on when the possible values of Y are the counting numbers $1,2,3,\ldots$. For example, we have:

  $$\Pr(S=0)=\Pr(N=0)$$

  $$\Pr(S=1)=\Pr(N=1 \text{ and } Y=1)=f_N(1)f_Y(1)$$

  $$\begin{aligned}\Pr(S=2)&=\Pr(N=2 \text{ and } Y_1=Y_2=1, \text{ or, } N=1 \text{ and } Y=2)\\&=f_N(2)f_Y(1)f_Y(1)+f_N(1)f_Y(2)\end{aligned}$$

In Chapters 3 and 4 we will see recursive ways to calculate these same probabilities when the frequency model is an $(a, b, 0)$ distribution. (These models are introduced in Chapter 3.) The starting value for this recursion is $\Pr(S=0)$. Calculation of this starting value is given by property (v) in the above theorem.

Example 1.10

Suppose that the annual frequency of losses from a portfolio follows a Poisson distribution with parameter $\lambda=10$:

$$\Pr(N=k) = e^{-10}\frac{10^k}{k!} \text{ for } k=0,1,2,\ldots$$
$$E[N]=\text{var}(N)=\lambda=10$$

Suppose that the individual loss amounts Y are uniformly distributed on $[0, 1000]$. Compute the mean and variance of aggregate annual losses against the portfolio.

Solution

To employ Properties (i) and (ii) of Theorem 1.2, we first need to calculate the mean and variance of the frequency and severity components:

$$E[N]=\text{var}(N)=\lambda=10 \quad \text{(given)}$$
$$f_Y(y) = 0.001 \quad \text{for } 0 < y < 1{,}000$$

Therefore:

$$E\left[Y^k\right] = \int_0^{1{,}000} y^k\, 0.001\, dy = \frac{1{,}000^{k+1}}{1{,}000(k+1)} = \frac{1{,}000^k}{k+1}$$
$$\Rightarrow \quad E[Y] = 500 \ , \ E\left[Y^2\right] = \frac{1{,}000{,}000}{3} \ , \ \text{var}(Y) = 83{,}333.33$$

From Property (i) of Theorem 1.2, the expected aggregate annual losses are:

$$E[S]=E[N]E[Y] = 10\times 500 = 5{,}000$$

By Property (ii) of Theorem 1.2, the variance of aggregate annual losses is:

$$\text{var}(S)=E[N]\ \text{var}(Y) + \left(E[Y]\right)^2 \text{var}(N)$$
$$= 10\times 83{,}333.33 + 500^2\times 10 = 3{,}333{,}333$$

♦♦

1.6 Mixing distributions to form new ones

The technique introduced here of combining several distributions to form a new distribution is called **mixing**. The distribution resulting from mixing is a hybrid of several other distributions.

Mixed distributions

You may be familiar with the most basic type of mixed distribution from a first course in probability. It has both a discrete part and a continuous part. These two parts can be determined from the CDF.

Here is an example from loss model theory of how a mixed distribution arises in a natural way.

Example 1.11

Suppose that a ground up loss X is uniformly distributed on $(0,1000]$. For each such loss suffered by a policyholder, suppose that an insurer will make a payment Y that is equal to the excess of the loss over 100 (provided that the loss exceeds 100) up to a maximum reimbursement of 500.

Determine the cumulative distribution function for Y, the insurance payment per loss and draw the graph of the CDF.

Solution

The first step is to write a formula for Y in terms of X:

$$Y = \begin{cases} 0 & \text{if } 0 < X \le 100 \\ X - 100 & \text{if } 100 < X < 600 \\ 500 & \text{if } 600 \le X \le 1,000 \end{cases}$$

Since the loss X is uniformly distributed, we have:

$$f_X(x) = 0.001 \quad \text{for } 0 < x \le 1,000$$

Now we can compute the CDF for Y:

- Since the loss X is positive, we have:

  $$F_Y(y) = 0 \ \text{ for } y < 0$$

- The event $Y = 0$ is equivalent to the event $0 < X \le 100$. So there is a point mass of probability at zero:

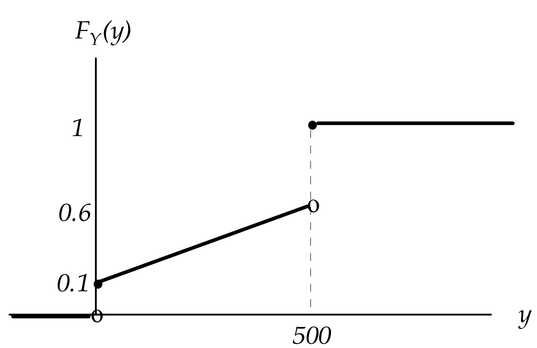
  $$F_Y(0) = \Pr(Y = 0) = \Pr(0 < X \le 100) = \int_0^{100} 0.001\,dx = 0.10$$

- The event $100 < X < 600$ is equivalent to the event $0 < Y < 500$. For $0 < y < 500$, we have:

  $$\begin{aligned} F_Y(y) &= \Pr(Y \le y) = \Pr(Y = 0) + \Pr(0 < Y \le y) \\ &= 0.10 + \Pr(0 < X - 100 \le y) = 0.10 + \Pr(100 < X \le y + 100) \\ &= 0.10 + \int_{100}^{y+100} 0.001\,dx = 0.10 + 0.001y \quad \text{for } 0 < y < 500 \end{aligned}$$

- $600 \le X \le 1,000 \Leftrightarrow Y = 500$. There is also a point mass of probability at 500.

The graph of the CDF is given below:



♦♦

Properties of a mixed distribution

1. There are point masses of probability at points y where $F_Y(y)$ has a jump discontinuity:

$$\Pr(Y=y) = F_Y(y) - \lim_{x \to y^-} F_Y(x) \qquad \text{(the height of the jump)}$$

This is the **discrete part** of the mixture**.**

2. The **continuous part** has a PDF equal to $F_Y'(y)$, where it exists.

3. Moments of the mixed distribution are computed as follows:

$$E\left[Y^k\right] = \sum_{\substack{\text{all } y \text{ where} \\ \Pr(Y=y)\neq 0}} y^k \Pr(Y=y) + \int_{-\infty}^{\infty} y^k F_Y'(y)\,dy$$

4. Probabilities of interval events are computed as follows:

$$\begin{aligned}
\Pr(a < Y \le b) &= F_Y(b) - F_Y(a) \\
\Pr(a \le Y \le b) &= \Pr(Y=a) + \Pr(a < Y \le b) \\
&= \left(F_Y(a) - \lim_{x \to a^-} F_Y(x)\right) + \left(F_Y(b) - F_Y(a)\right) \\
&= F_Y(b) - \lim_{x \to a^-} F_Y(x)
\end{aligned}$$

Example 1.12

Compute the mean and variance of Y in Example 1.11.

Solution

We will use Property 3 since Y has a mixed distribution. The discrete part has point masses of probability at 0 and 500. The respective heights of the jump discontinuities at these points are 0.10 and 0.40. The continuous part has a PDF that is non-zero on the interval $(0, 500)$. For $0 < y < 500$, we saw that $F_Y(y) = 0.10 + 0.001y$ in Example 1.11. Looking back at that formula and the accompanying graph, you can see that the derivative fails to exist at 0 and 500, the derivative is zero for $y < 0$ and for $y > 500$, and the derivative is equal to 0.001 for $0 < y < 500$.

Here is the formula for $f_Y(y)$:

$$f_Y(y) = \begin{cases} 0.10 & y = 0 & \text{(discrete part)} \\ 0.40 & y = 500 & \text{(discrete part)} \\ 0.001 & 0 < y < 500 & \text{(continuous part)} \end{cases}$$

So the moments are computed as follows:

$$E\left[Y^k\right] = 0^k \times 0.10 + 500^k \times 0.40 + \int_0^{500} y^k\, 0.001\, dy$$

Therefore:

$$E[Y] = 200 + 125 = 325$$
$$E\left[Y^2\right] = 100{,}000 + 41{,}666.67 = 141{,}666.67$$

Hence the variance is:

$$\text{var}(Y) = E\left[Y^2\right] - \left(E[Y]\right)^2 = 36{,}041.67 \qquad \blacklozenge\blacklozenge$$

There are several other ways to make these same calculations. You actually encountered some mixed distributions in Exam M.

For example, if Z is the random present value variable for a 10-year term insurance of 1 on (x) with the benefit payable on death (a continuous model), then Z is a function of $T(x)$ (future lifetime). It has a point mass of probability at $Z = 0$ corresponding to the event $T(x) > 10$, since in this case no payment is made. The continuous part of the distribution of Z corresponds to the event $0 < T(x) \le 10$. You can avoid the problem of dealing with Z as a mixed distribution by computing moments and probabilities in terms of the distribution of $T(x)$.

Example 1.13

Compute the first and second moment of Y in Example 1.11 by viewing Y as a function of the ground up loss X.

Solution

$$E\left[Y^k\right] = \int_0^{100} 0^k f_X(x)\,dx + \int_{100}^{600} (x-100)^k f_X(x)\,dx + \int_{600}^{1000} 500^k f_X(x)\,dx$$

$$\Rightarrow \quad E[Y] = 0 + 125 + 200 = 325 \quad , \quad E\left[Y^2\right] = 0 + 41{,}666.67 + 100{,}000 = 141{,}666.67 \qquad \blacklozenge\blacklozenge$$

One other thing can be done with the PDF of a mixed distribution. It can be written as a **weighted average** of a discrete probability function and a continuous PDF. For the payment per loss variable Y in Examples 1.11 - 1.13, we saw that:

$$f_Y(y) = \begin{cases} 0.10 & y = 0 & \text{(discrete part)} \\ 0.40 & y = 500 & \text{(discrete part)} \\ 0.001 & 0 < y < 500 & \text{(continuous part)} \\ 0 & \text{otherwise} \end{cases}$$

Consider a discrete random variable D with point masses of probability at 0 and 500 proportional to 0.10 and 0.40 respectively:

$$f_D(0) = \frac{0.10}{0.50} = 0.20 \qquad f_D(500) = \frac{0.40}{0.50} = 0.80 \quad \text{and zero otherwise}$$

Consider a continuous random variable C whose PDF is proportional to the continuous part of f_Y:

$$f_C(y) = \frac{0.001}{0.500} = 0.002 \ \text{ for } \ 0 < y < 500 \quad \text{and zero otherwise}$$

Then we can rewrite $f_Y(y)$ as follows:

$$f_Y(y) = 0.50\ f_D(y) + 0.50\ f_C(y)$$

The two coefficients equal to 0.50 are the **weights**. Moments of Y can now be computed in a third way from this weighted average:

$$E\left[Y^k\right] = 0.50\ E\left[D^k\right] + 0.50\,E\left[C^k\right]$$

Here is why that will work. In the solution to Example 1.12 we saw the following:

$$E\left[Y^k\right] = 0^k \times 0.10 + 500^k \times 0.40 + \int_0^{500} y^k\, 0.001\, dy$$

Now we will rearrange this moment formula slightly:

$$\begin{aligned} E\left[Y^k\right] &= 0^k \times 0.10 + 500^k \times 0.40 + \int_0^{500} y^k\, 0.001\, dy \\ &= 0.50 \underbrace{\left(0^k \times \frac{0.10}{0.50} + 500^k \times \frac{0.40}{0.50}\right)}_{E\left[D^k\right]} + 0.50 \underbrace{\int_0^{500} y^k\, \frac{0.001}{0.50}\, dy}_{E\left[C^k\right]} \end{aligned}$$

The reason that we went through this discussion of a weighted average approach is that this is the model for the more general types of mixing that we will consider next.

Note at this point that it is *not* true to say that:

$$\text{var}(Y) = 0.50\,\text{var}(C) + 0.50\,\text{var}(D)$$

In general the variance of a mixed distribution cannot be expressed as the weighted average of the individual component variances.

Mixing with a discrete parameter

Let's start with an example to introduce the theory.

As you might know, the Poisson distribution (a discrete counting distribution) is often used to model the frequency with which some type of random event occurs over some fixed time period. Suppose that we consider an insurer's portfolio of auto insurance policies for the year 2008. At the start of that year we might assume that the random number of claims to be filed for losses in year 2008 against a policy, N, might follow a Poisson distribution.

So we would have:

$$\Pr(N = k) = e^{-\lambda}\frac{\lambda^k}{k!} \quad \text{for } k = 0, 1, \ldots \quad \text{where } \lambda > 0$$

$$E[N] = \text{var}(N) = \lambda$$

But, individuals drive different distances over a year, they live in different areas with different hazards due to traffic patterns, and so on. So not all drivers and policies have the same characteristics. Therefore the insurer might use different values of λ for different policies.

Suppose the insurer used past data to classify current policyholders as follows:

Category	**Parameter**	**Percentage of policies**
low risk	$\lambda=0.20$	35%
moderate risk	$\lambda=0.40$	60%
high risk	$\lambda=1.00$	5%

Since the value of λ varies over the portfolio, we should consider it as a discrete random variable Λ with probability function:

$$f_\Lambda(0.20)=0.35 \quad , \quad f_\Lambda(0.40)=0.60 \quad , \quad f_\Lambda(1.00)=0.05$$

In this light, the probability function for annual claim frequency for a policy should be viewed as a *conditional* Poisson distribution:

$$\Pr(N=k \mid \Lambda=\lambda) = e^{-\lambda}\frac{\lambda^k}{k!} \quad \text{for } k=0,1,...$$

So we have changed our point of view. Instead of viewing the parameter as a number, we now view it as a random variable.

Now suppose that we have a new policyholder at the start of year 2008 with no past data to help us classify the risk level. The policyholder will have to pass underwriting, but this person is a randomly selected policy from the portfolio. So the appropriate annual claims frequency N should be modeled as the marginal distribution where the effect of λ has been summed out.

Here's how we obtain the marginal distribution of N:

- The first step is to multiply the marginal probability function of Λ and the conditional probability function of N given Λ to obtain the *joint probability function* of N and Λ:

$$f_{N,\Lambda}(k,\lambda)=f_\Lambda(\lambda)\,f_N(k|\Lambda=\lambda)$$

- The second step is to sum this joint probability function over all values of λ to obtain the *marginal probability function* of N:

$$f_N(k)=\sum_\lambda f_{N,\Lambda}(k,\lambda)=\sum_\lambda f_\Lambda(\lambda)\,f_N(k \mid \Lambda=\lambda)$$

$$= 0.35\,e^{-0.2}\frac{0.2^k}{k!} + 0.60\,e^{-0.4}\frac{0.4^k}{k!} + 0.05\,e^{-1.0}\frac{1.0^k}{k!} \quad \text{for } k=0,1,...$$

This expression could also be referred to as a **weighted average of Poisson distributions,** where 0.35, 0.60, and 0.05 are the **weights** that sum to 1. Once this expression is written down it is easy to calculate probabilities or moments for N.

Example 1.14

For the portfolio of auto insurances described above, compute:

(a) The probability that there are 2 or more claims in the next year for a new policyholder.

(b) The expected number of claims for a randomly selected policyholder in the next year.

Solution

(a) The probability that is requested is computed as follows. From the marginal probability function above, we can calculate:

$$f_N(0) = 0.70714 \;\;,\;\; f_N(1) = 0.23658$$

Therefore:

$$\begin{aligned}\Pr(N \geq 2) &= 1 - 0.70714 - 0.23658 \\ &= 0.05628\end{aligned}$$

(b) The expected value is computed as follows:

$$\begin{aligned}E[N] &= \sum_k k \Pr(N = k) \\ &= \sum_{k=0}^{\infty} k \left(0.35\, e^{-0.2} \frac{0.2^k}{k!} + 0.60\, e^{-0.4} \frac{0.4^k}{k!} + 0.05\, e^{-1.0} \frac{1.0^k}{k!} \right) \\ &= 0.35 \underbrace{\sum_{k=0}^{\infty} k\, e^{-0.2} \frac{0.2^k}{k!}}_{E[N|\Lambda=0.2]=0.2} + 0.60 \underbrace{\sum_{k=0}^{\infty} k\, e^{-0.4} \frac{0.4^k}{k!}}_{E[N|\Lambda=0.4]=0.4} + 0.05 \underbrace{\sum_{k=0}^{\infty} k\, e^{-1.0} \frac{1.0^k}{k!}}_{E[N|\Lambda=1.0]=1.0} \\ &= 0.36\end{aligned}$$

Note: You can see from the last line that this calculation is equivalent to using the Double Expectation Theorem: $E[N] = E\big[E[N \mid \Lambda] \big]$. ♦♦

Mixing with a discrete parameter

You are given $f_X(x \mid \Theta = \theta)$ where Θ is a discrete random variable.

1. The joint PDF of X and Θ is:

$$f_{X,\Theta}(x,\theta) = f_\Theta(\theta)\, f_X(x \mid \Theta = \theta)$$

2. The marginal PDF of X is computed as follows:

$$f_X(x) = \sum_\theta f_{X,\Theta}(x,\theta) = \sum_\theta f_\Theta(\theta)\, f_X(x \mid \Theta = \theta)$$

This is a weighted average of $f_X(x \mid \Theta = \theta)$.

3. Moments of X are computed as weighted averages of conditional moments as follows:

$$E\left[X^k\right] = E\left[E\left[X^k \mid \Theta \right]\right] = \sum_\theta \Pr(\Theta = \theta) E\left[X^k \mid \Theta = \theta \right]$$

Two-point mixture

When the mixing parameter Θ has only 2 possible values, then the distribution of X is called a **two-point mixture**.

Take another look at Example 1.9 (Section 1.4) where we had a population of lives that was a mixture of males and females with slightly different mortality models. The mortality model for a randomly selected life from this population is a two-point mixture. Try to obtain the results of this example using the theory outlined above.

Mixing with a continuous parameter

The theory here is virtually identical to the theory of mixing with a discrete parameter. Where you summed in the case of a discrete distribution for the parameter, here you will integrate.

Mixing with a continuous parameter

You are given $f_X(x|\Theta=\theta)$ where Θ is a continuous random variable.

1. The joint PDF of X and Θ is:

$$f_{X,\Theta}(x,\theta) = f_\Theta(\theta)\, f_X(x \mid \Theta=\theta)$$

2. The marginal PDF of X is computed as follows:

$$f_X(x) = \int_\theta f_{X,\Theta}(x,\theta)\,d\theta = \int_\theta f_\Theta(\theta)\, f_X(x \mid \Theta=\theta)\,d\theta$$

3. Moments of X are computed as follows:

$$E\left[X^k\right] = E\left[E\left[X^k \mid \Theta\right]\right] = \int_\theta f_\Theta(\theta) E\left[X^k \mid \Theta=\theta\right] d\theta$$

Example 1.15

Suppose that you are given:

$$f_N(k|\Theta=\theta) = e^{-\theta}\frac{\theta^k}{k!} \quad \text{for } k=0,1,\ldots$$

where: $f_\Theta(\theta) = \dfrac{1}{0.8}$ for $0.2 \le \theta \le 1.0$

Determine $\Pr(N \ge 1)$ and $E[N]$.

Solution

We again have a conditional Poisson distribution for frequency. But the parameter is distributed uniformly on the interval $[0.2, 1.0]$.

We know that $\Pr(N \ge 1) = 1 - f_N(0)$. From Property 2 we have:

$$f_N(0) = \int f_{N,\Theta}(0,\theta)\,d\theta = \int f_\Theta(\theta)\, f_N(0 \mid \Theta=\theta)\,d\theta$$

$$= \int_{0.2}^{1.0} \frac{1}{0.8} e^{-\theta}\,d\theta = \frac{e^{-0.2} - e^{-1.0}}{0.8} = 0.56356$$

So we have:

$$\Pr(N \ge 1) = 1 - f_N(0) = 0.43644$$

Furthermore, since $E[N|\Theta=\theta]=\theta$ (the conditional distribution is Poisson with parameter θ), we have:

$$E[N]=E\left[E[N \mid \Theta]\right] = E[\Theta] = \frac{0.2+1.0}{2} = 0.6 \quad \text{(mean of a uniform distribution)}$$ ♦♦

Example 1.16

Each life in a certain population has a lifetime that follows a constant force model (Exam M):

$$f_X(x|M=\mu) = \mu e^{-\mu x} \quad \text{for} \quad 0 < x < \infty$$

For this population, the force varies uniformly from 1.0 to 2.0.

Determine the probability that a randomly selected life survives 1 year.

Solution

We are asked to compute the unconditional probability $s_X(1) = \Pr(X > 1)$, since for a "randomly selected life" we do not know (*ie* are not given) the appropriate force.

We are given that:

$$f_M(\mu) = 1 \quad \text{for} \quad 1 < \mu < 2$$

To compute the marginal PDF for X we first need the joint PDF for X and M:

$$f_{X,M}(x,\mu) = \mu e^{-\mu x} \cdot 1 \quad \text{for} \quad 0 < x < \infty \text{ and } 1 < \mu < 2$$

Now integrate out the variable μ:

$$f_X(x) = \int_{\mu=1}^{2} f_{X,M}(x,\mu)\,d\mu = \int_{\mu=1}^{2} \mu e^{-\mu x}\,d\mu$$

Let's pause briefly to think about what is needed to finish the problem. Evaluating this integral is the tricky part because you now look at x as a constant and you have to anti-differentiate with respect to μ. So looking at the integral above, you see that integration by parts would be required. Furthermore, even after this integration is completed, you will still need to perform another integral calculation:

$$s_X(1) = \int_{1}^{\infty} f_X(x)\,dx$$

There is a possible way to avoid these complications with a bit of theory we have not yet developed. In the same manner that the marginal PDF was obtained as a weighted average of the conditional PDF, you can also obtain the marginal survival function as a weighted average of the conditional survival function:

$$s_X(x) = \int_{\mu=1}^{2} s_X(x \mid M=\mu) f_M(\mu)\,d\mu$$

This integral is considerably simpler than the one for the marginal PDF because $s_X(x \mid M=\mu) = e^{-\mu x}$. It also avoids the problem of an additional integral of the PDF to obtain the survival function.

We can now get there in one step:

$$s_X(x) = \int_{\mu=1}^{2} s_X(x \mid M=\mu) f_M(\mu)\,d\mu = \int_{\mu=1}^{2} e^{-\mu x} \cdot 1\,d\mu$$

$$= \left(-\frac{e^{-\mu x}}{x} \right)\Bigg|_{\mu=1}^{2} = \frac{e^{-x} - e^{-2x}}{x}$$

$$\Rightarrow \quad s_X(1) = 0.23254$$

♦♦

Theorem 1.3

(i) Suppose that $X|\Theta \sim$ exponential mean Θ and $\Theta \sim$ inverse gamma distribution with parameters α, θ. Then X follows a 2-parameter Pareto distribution with parameters α, θ matching the parameters of the inverse gamma mixing distribution.

(ii) Suppose that $X|\Theta \sim \text{Normal}\left(\Theta, \sigma_1^2\right)$ and Θ follows the normal distribution $\text{Normal}\left(\mu, \sigma_2^2\right)$. Then X follows the normal distribution $\text{Normal}\left(\mu,\, \sigma_1^2 + \sigma_2^2\right)$.

Proof

(i) Here are the details for part (i). We will use λ to denote the mean of the exponential distribution instead of the usual θ since θ is a parameter of the inverse gamma distribution.

1. $f_X(x|\Lambda = \lambda) = \dfrac{e^{-x/\lambda}}{\lambda}$ for $x > 0$ and $\lambda > 0$ (conditional exponential)

2. $f_\Lambda(\lambda) = \dfrac{\theta^\alpha e^{-\theta/\lambda}}{\lambda^{\alpha+1}\Gamma(\alpha)}$ for $\lambda > 0$ (marginal inverse gamma: see exam tables)

3. $$f_{X,\Lambda}(x,\lambda) = f_\Lambda(\lambda) f_X(x|\Lambda = \lambda) = \frac{\theta^\alpha e^{-(\theta+x)/\lambda}}{\lambda^{\alpha+2}\Gamma(\alpha)}$$
$$= \frac{\alpha\theta^\alpha}{(\theta+x)^{\alpha+1}} \times \frac{(\theta+x)^{\alpha+1} e^{-(\theta+x)/\lambda}}{\lambda^{\alpha+2}\Gamma(\alpha+1)} \quad (\Gamma(\alpha+1) = \alpha\Gamma(\alpha))$$

4. $$f_X(x) = \int_0^\infty f_{X,\Lambda}(x,\lambda)\, d\lambda$$
$$= \frac{\alpha\theta^\alpha}{(\theta+x)^{\alpha+1}} \int_0^\infty \frac{(\theta+x)^{\alpha+1} e^{-(\theta+x)/\lambda}}{\lambda^{\alpha+2}\Gamma(\alpha+1)}\, d\lambda$$
$$= \frac{\alpha\theta^\alpha}{(\theta+x)^{\alpha+1}} \quad \text{(the 2-parameter Pareto PDF)}$$

(ii) The same four steps as in the proof of the first part should be repeated. But this time the algebra in Step 4 is much more complex. So we will not include a full proof. Instead we will simply use the Double Expectation Theorem to show that the marginal mean and variance agree with the assertion in the Theorem:

1. $X|\Theta \sim \text{Normal}\left(\Theta, \sigma_1^2\right) \Rightarrow$
$E\left[X|\Theta\right] = \Theta$, $\text{var}(X|\Theta) = \sigma_1^2$

2. $E[X] = E\left[E\left[X|\Theta\right]\right] = E[\Theta] = \mu$ $\quad (\Theta \sim \text{Normal}\left(\mu, \sigma_2^2\right))$

3. $$\text{var}(X) = E\left[\text{var}(X|\Theta)\right] + \text{var}\left(E\left[X \mid \Theta\right]\right)$$
$$= E\left[\sigma_1^2\right] + \text{var}(\Theta) = \sigma_1^2 + \sigma_2^2$$

□

1.7 An example with the IRM and CRM models

Suppose that an insurer has a portfolio of 100 independent policies. Assume that over the next year each policy will generate either 0 or 1 claim with respective probabilities 0.90 and 0.10. Assume that each claim amount is uniformly distributed on $(0,1000]$. Then we have:

$$\Pr(N=0)=0.90$$
$$\Pr(N=1)=0.10$$
$$f_Y(y)=0.001 \text{ for } 0<y<1000$$

Suppose we let S denote the insurer's aggregate claims in the next year. Let's compute the expected value and variance of S using both the IRM and CRM models. You will see that mixed distributions play a role in this analysis. This example will serve as a reminder of the properties of these two methods of modeling and it should be read several times until it is completely understood.

IRM

Let's proceed first according to the IRM. Aggregate annual claims are modeled by:

$$S=X_1+\cdots+X_{100}$$

where X_i is the total annual claims arising from the ith policy, and the various X_i are independent and identically distributed like X.

What is the distribution of X? Since each policy will experience 0 or 1 claim, the total annual claim amount is either 0 with probability 0.90, or it is uniformly distributed on $(0,1000]$ with probability 0.10. This is a two-point mixture:

$$f_X(x)=\begin{cases} 0.90 & \text{if } x=0 \quad \text{(discrete part)} \\ 0.0001 & \text{if } 0<x<1000 \quad \text{(continuous part)} \end{cases}$$

So the mean and variance of X are computed as follows:

$$E\left[X^k\right]=0^k\times 0.90+\left(\int_0^{1000} x^k\, 0.0001\, dx\right)$$
$$\Rightarrow \quad E[X]=50\,,\quad E\left[X^2\right]=33{,}333.33$$
$$\Rightarrow \quad \operatorname{var}(X)=30{,}833.33$$

Finally, from the IRM equation $S=X_1+\cdots+X_{100}$, it follows that we have:

$$E[S]=100\,E[X]=5{,}000$$
$$\operatorname{var}(S)=100\operatorname{var}(X)=3{,}083{,}333.33$$

CRM

Now lets apply the CRM. In this model we view aggregate annual claims as:

$$S=Y_1+\cdots+Y_N$$

where Y is the individual claim amount model (uniform on $(0,1000]$), and N is the annual frequency of claims arising from the whole portfolio.

The claim frequency for a single policy follows a Bernoulli distribution with $p=0.10$. So the annual claim frequency for the whole portfolio is a sum of 100 independent Bernoulli variables with $p=0.10$. In other words, the distribution of N is binomial with $n=100$, $p=0.10$.

Recall from Theorem 1.2 (Section 1.5) that the mean and variance for a CRM model are computed as:

$$E[S]=E[N]E[Y] \qquad \operatorname{var}(S)=E[N]\operatorname{var}(Y) + \left(E[Y]\right)^2 \operatorname{var}(N)$$

Since the severity model is uniform on (0,1000], we have:

$$E[Y] = \frac{1{,}000}{2} = 500 \qquad \operatorname{var}(Y) = \frac{1{,}000^2}{12} = 83{,}333.33$$

The frequency model is binomial, so we have:

$$E[N] = np = 100 \times 0.10 = 10$$
$$\operatorname{var}(N) = np(1-p) = 9$$

As a result, for the aggregate annual claims we have:

$$E[S]=E[N]E[Y]$$
$$= 10 \times 500 = 5{,}000$$
$$\operatorname{var}(S)=E[N]\operatorname{var}(Y) + \left(E[Y]\right)^2 \operatorname{var}(N)$$
$$= 10 \times 83{,}333.33 + (500)^2 \times 9 = 3{,}083{,}333.33$$

1.8 Transformations of random variables

Suppose that X is a continuous random variable, and that $Y = g(X)$ is a differentiable function of X that is 1-1. We want to describe the relation between the two PDF's f_X and f_Y for certain commonly used transformations.

You should already be familiar with most of the results here from a basic probability course. If you need further explanation, you should consult a probability text.

According to the **method of transformations**, the relation between f_X and f_Y is:

$$f_Y(y) = f_X\left(g^{-1}(y)\right)\left|\frac{d\,g^{-1}(y)}{dy}\right| \quad \text{where } g^{-1}(y) \text{ is the inverse function}$$

There are basically two types of 1-1, differentiable transformation: increasing and decreasing. The inverse will exist in each case.

If $g(x)$ is an increasing function then so is the inverse function. Similarly, if $g(x)$ is a decreasing function, then so is the inverse function. So the absolute value in the formula above is essential when the transformation is decreasing and has a negative derivative.

The particular transformations that are frequently used in loss models are:

Linear:	$Y = aX$	models an inflationary effect
Exponential:	$Y = e^X$	X normally distributed $\Rightarrow Y$ is lognormal
Raise to a power:	$Y = X^\gamma$	X exponentially distributed
		eg X^{-1} is inverse exponential

In each case we have 1-1, differentiable functions, so the method of transformations can be applied to develop a formula for f_Y.

Example 1.17

Suppose that X follows a gamma distribution (see Chapter 2 for more details) with parameters θ and α then:

$$f(x)=\frac{1}{\theta^{\alpha}\Gamma(\alpha)}x^{\alpha-1}e^{-x/\theta} \quad \text{where } 0<x<\infty \text{ and } \theta>0,\ \alpha>0$$

Determine the distribution of $Y=aX$ where $a>0$.

Solution

For the linear transformation $Y=g(X)=aX$, the inverse transformation is $X=g^{-1}(Y)=Y/a$. So by the method of transformations, we have:

$$\begin{aligned}f_Y(y)&=f_X\left(g^{-1}(y)\right)\left|\frac{dg^{-1}(y)}{dy}\right|=f_X(y/a)/|a|\\&=\frac{1}{\theta^{\alpha}\Gamma(\alpha)}(y/a)^{\alpha-1}e^{-(y/a)/\theta}/a \qquad \text{(given } a>0\text{)}\\&=\frac{1}{(a\theta)^{\alpha}\Gamma(\alpha)}y^{\alpha-1}e^{-y/(a\theta)}\end{aligned}$$

If you look carefully at this relation you should see that Y follows a gamma distribution with the same α parameter as X, but with the parameter θ being replaced with $a\theta$. ♦♦

Note that you can also prove this result by considering the moment generating function of the gamma distribution. You may like to check that you can obtain the same result using this alternative method.

Certain parametric families of distributions that will be used for severity models exhibit the property: if X is in the family, then so is aX. Such a family is called a **scale family.** In the preceding example we saw that the gamma family is a scale family. Another example is the normal family:

$$X\sim N\left(\mu,\sigma^2\right) \Rightarrow aX\sim N\left(a\mu,a^2\sigma^2\right)$$

Sometimes a scale family has a **scale parameter.** A parameter of X is called a scale parameter if the corresponding parameter of aX is multiplied by a, and all other parameters are unchanged. For example, we showed in Example 1.17 that the parameter θ in the gamma distribution is a scale parameter. In contrast, the normal family does not have a scale parameter. Both the mean and variance parameters are changed when X is multiplied by a.

1.9 Splicing distributions together

Another technique for combining several continuous models to form a hybrid of the components is called **splicing**.

If X is a continuous loss model you might want to use different distributions for different ranges of losses. For example suppose we want to create a model so that losses in the range $(0,1000]$ are uniformly distributed and losses in the range $(1000,3000]$ are also uniformly distributed. Suppose that 85% of losses are less than or equal to 1,000.

Splicing these two distributions together is accomplished as follows:

$$f_1(x) = 0.001 \qquad \text{for } 0 < x \le 1{,}000 \qquad \text{(uniform on } (0, 1000]$$
$$f_2(x) = 0.0005 \qquad \text{for } 1{,}000 < x \le 3{,}000 \qquad \text{(uniform on } (1000, 3000])$$
$$f(x) = \begin{cases} 0.85\, f_1(x) & \text{for } 0 < x \le 1{,}000 \\ 0.15\, f_2(x) & \text{for } 1000 < x \le 3{,}000 \end{cases} \qquad \text{(the spliced distribution)}$$

Notice that this is really just the PDF of a continuous model:

$$\int_{-\infty}^{\infty} f(x)\,dx = 0.85 \int_0^{1{,}000} 0.001\,dx + 0.15 \int_{1{,}000}^{3{,}000} 0.0005\,dx$$
$$= 0.85\times 1 + 0.15\times 1 = 1.0$$

The **most general type of splicing** of n different distributions together is described as follows:

- For $i = 1, 2, \ldots, n$, you are given a continuous model PDF $f_i(x)$ with all probability on $(a_i, a_{i+1}]$ where $0 \le a_1 < a_2 < \cdots < a_{n+1}$
- For $i = 1, 2, \ldots, n$, you are given a probability p_i so that $p_1 + p_2 + \cdots + p_n = 1$
- The spliced PDF is defined by:

  $$f(x) = p_i\, f_i(x) \quad \text{for } a_i < x \le a_{i+1} \text{ and } i = 1, \ldots, n$$

Example 1.18

Suppose that a new loss distribution X is created by splicing the uniform distribution on $(0, 1000]$ with the uniform distribution on $(1000, 3000]$ where $p_1 = 0.85$ and $p_2 = 0.15$.

Determine $\Pr(X \le 2{,}000)$ and $E[X]$.

Solution

This is the model discussed at the beginning of this section. We have seen that:

$$f(x) = \begin{cases} 0.85 \times 0.001 & \text{for } 0 < x \le 1{,}000 \\ 0.15 \times 0.0005 & \text{for } 1000 < x \le 3{,}000 \end{cases}$$

As a result we have:

$$\Pr(X \le 2{,}000) = \int_0^{1{,}000} 0.85 \times 0.001\,dx + \int_{1{,}000}^{2{,}000} 0.15 \times 0.0005\,dx$$
$$= 0.85 + 0.075 = 0.925$$

$$E[X] = \int_{a_1}^{a_3} x\, f(x)\,dx = p_1 \int_{a_1}^{a_2} x\, f_1(x)\,dx + p_2 \int_{a_2}^{a_3} x\, f_2(x)\,dx \quad \text{(general)}$$
$$= 0.85 \int_0^{1{,}000} 0.001 x\,dx + 0.15 \int_{1{,}000}^{3{,}000} 0.0005\, x\,dx = 725$$

♦♦

1.10 Truncation and censoring of distributions

Distributions can be **truncated** or **censored** when there is a loss of information.

Truncation

As an example of truncation, consider the case where there is a deductible operating on an insurance policy (covered in detail in Chapter 2). For example, suppose that for each ground up loss X greater than 100, an insurer pays a claim Y, the excess of X over 100. So we have:

$$Y = (X - 100) \mid X > 100 = (X \mid X > 100) - 100$$

The distributions of Y and X are closely related, but the only values of X that the insurer is aware of are ones greater than 100. If a loss X is less than or equal to 100, there is no claim to be filed and paid by the insurer. So for some X's there is no Y. There is a loss of information as you move from X to Y. This is an example of **truncation below** at 100. In general, any variable Y that is a conditional form of X, given that X falls in some interval, is obtained by **truncating** the distribution of X. For some values of X, there will be no value of Y to observe.

We have already met this phenomenon in survival model theory. The future lifetime after age x is obtained by truncating and shifting the distribution of the lifetime of a newborn:

$$T(x) = (X - x) \mid X > x = (X \mid X > x) - x$$

The main result that is needed when dealing with a random variable Y that is obtained by truncating X is a method for finding the PDF of Y from the PDF of X. This involves a simple argument with conditional probabilities.

Suppose that $Y = X \mid X \in I$ for some interval $I = (a, b)$. We will first derive the relationship between the two distribution functions.

For $x \in (a, b)$, we have:

$$F_Y(x) = \Pr(Y \le x) = \Pr(X \le x \mid X \in (a, b)) = \frac{\Pr(X \le x \text{ and } a < X < b)}{\Pr(a < X < b)}$$

$$= \frac{\Pr(a < X \le x)}{\Pr(a < X < b)} \qquad \text{(since } x < b\text{)}$$

$$= \frac{F_X(x) - F_X(a)}{\Pr(a < X < b)} \qquad \text{for } a < x < b$$

This relation can be used when X is either discrete or continuous.

When X is a *continuous* random variable, the denominator above is the same as $F_X(b) - F_X(a)$, and you can differentiate this relation to relate the PDF of Y to the PDF of X:

$$f_Y(x) = F_Y'(x) = \frac{f_X(x)}{\Pr(a < X < b)} = \frac{f_X(x)}{F_X(b) - F_X(a)} \qquad \text{for } a < x < b$$

In the *discrete* case we also have $f_Y(x) = \dfrac{f_X(x)}{\Pr(a < X < b)}$ for $a < x < b$.

Example 1.19

Suppose that $l_x = (80-x)^2$ for $0 \le x \le 80$ and let X be the corresponding lifetime of a newborn.

Determine the PDF of Y, the age at death of a newborn who is known to die after age 50.

Solution

From this description we have $Y = X | X > 50$. So the PDF of Y is:

$$f_Y(x) = \frac{f_X(x)}{F_X(80) - F_X(50)} \quad \text{for} \quad 50 < x \le 80$$

The first step is to calculate f_X:

$$s_X(x) = \frac{l_x}{l_0} = \frac{(80-x)^2}{80^2}$$

$$\Rightarrow f_X(x) = -s'_X(x) = \frac{2(80-x)}{80^2} \quad \text{for} \quad 0 < x \le 80$$

Finally, we have:

$$f_Y(x) = \frac{f_X(x)}{F_X(80) - F_X(50)} = \frac{2(80-x)/80^2}{1 - \left(1 - \frac{30^2}{80^2}\right)} = \frac{2(80-x)}{30^2} \quad \text{for} \quad 50 < x \le 80$$

♦♦

Censoring

Another important concept is censoring. If X is a random variable we defined $Y = X \wedge n$ as:

$$X \wedge n = \min\{X, n\} = \begin{cases} X & \text{if } X \le n \\ n & \text{if } X > n \end{cases}$$

The variable Y is a **censored form** of X. For every X there is a Y. However, if we observe $Y = n$, then we only know that $X \ge n$. The precise value of X is unknown. In other words, from looking at the value of Y sometimes the value of X is known precisely (when $Y < n$), but for other values of Y the value of X is only known to be in some interval (when $Y = n$).

Information is lost here in the sense that when you observe Y some X values are known imprecisely.

The temporary life expectancy $\mathring{e}_{x:\overline{n}|}$ may be defined as $E\left[T(x) \wedge n\right]$, the expected value of a censored form of the future lifetime at age x.

In loss models, a similar type of censoring occurs as follows. Suppose that X is the ground up loss of a policyholder, and suppose the insurance payment per loss, Y, is $X \wedge L$. In other words, losses less than X are fully reimbursed, but losses X that are bigger than L result in a reimbursement of L. The number L is known as a **policy limit** (discussed further in Chapter 2).

Example 1.20

Suppose that a ground up loss X is uniformly distributed on the interval $(0,1000]$. Suppose there is a policy limit of 500 per loss.

Determine the insurer's expected payment per loss.

Solution

We are given:

$$f_X(x) = 0.001 \text{ for } 0 < x \le 1{,}000 \text{ and } Y = X \wedge 500 = \min\{X, 500\}$$

The expected payment per loss is thus:

$$E[Y] = \int_0^{500} x\, f_X(x)\,dx + \int_{500}^{1{,}000} 500\, f_X(x)\,dx = \int_0^{500} 0.001x\,dx + \int_{500}^{1{,}000} 500 \times 0.001\,dx$$

$$= \frac{500^2 \times 0.001}{2} + 500^2 \times 0.001 = 125 + 250 = 375$$ ♦♦

It is interesting to note that when a continuous random variable X is censored above at L, that is, we observe $Y = X \wedge L$, the resulting variable has a mixed distribution. The PDF of Y is given by:

$$f_Y(x) = \begin{cases} f_X(x) & \text{if } x < L \text{ (continuous part)} \\ \Pr(X \ge L) & \text{if } x = L \text{ (discrete part)} \end{cases}$$

This is simply due to the fact that when $X = x < L$ we have $Y = x$ as well. And there is obviously a point mass probability at $Y = L$, since this event is equivalent to the event $X \ge L$ that has probability equal to $s_X(L)$.

Chapter 1 Practice Questions

Question 1.1

A continuous random variable X has a moment generating function given by $M_X(t) = (1-10t)^{-2}$ for $t < 0.10$. Determine the mean and variance of X.

Question 1.2

The probability function for a discrete random variable N is given by:

$$\Pr(N = k) = \binom{4}{k} 0.7^k \, 0.3^{4-k} \quad \text{for } k = 0,1,2,3,4$$

Use the binomial theorem to obtain a closed form formula for the moment generating function of N. Then use this formula to write down a formula for the probability generating function.

Question 1.3

If N_1 and N_2 are independent and identically distributed like N in Question 1.2, find the probability generating function for $N_1 + N_2$.

Question 1.4

In Example 1.6 we found $f_X^{*3}(x)$ where $f_X(1) = 0.7$ and $f_X(2) = 0.3$. Use the results of this Example along with the recursive method to calculate $f_X^{*4}(x)$.

Question 1.5

An insurer has a portfolio of 100 policies. For each policy there is an 80% chance that the total annual loss is zero. There is a 20% chance that the total annual loss is uniformly distributed on the interval $[100, 2{,}000]$. Determine the mean, variance, and approximate 95-th percentile for the insurer's aggregate annual losses. Use the Central Limit Theorem to approximate the distribution of aggregate annual losses.

Question 1.6

In a certain group of 23-year-olds the males have a mean height of 70 inches and a standard deviation of 4 inches. The females have a mean height of 67 inches with a standard deviation of 3 inches. Males make up 45% of the group. Find the mean and standard deviation in the height of a randomly selected member of this group.

Question 1.7

A portfolio consists of 4 policies. The individual loss amounts for a 1-year period and the policy numbers are as follows:

57 (policy #2), 100 (Policy #4), 90 (Policy #2), 140 (Policy #1), 30 (Policy #1)

Using the notation of Section 1.5, determine X_1, X_2, X_3, X_4 and S. Then determine the values of $N, Y_1, Y_2, \ldots, Y_N$ and S.

Question 1.8

The annual frequency of losses from a portfolio, N, follows a Poisson distribution with mean $\lambda = 5$:

$$\Pr(N = k) = e^{-5}\frac{5^k}{k!} \quad k = 0, 1, 2, \ldots$$

The individual loss amounts, Y, follow an exponential distribution with PDF:

$$f_Y(y) = 0.002\, e^{-0.002y} \quad \text{for } y > 0$$

Let S be the aggregate annual loss for the portfolio.

Determine $E[S]$, $\text{var}(S)$, $M_S(t)$, and $\Pr(S = 0)$.

Question 1.9

A ground up loss follows the PDF:

$$f_X(x) = \frac{2(1{,}000 - x)}{1{,}000^2} \quad \text{for } 0 \le x \le 1{,}000$$

For each such loss by a policyholder an insurer pays Y. Y is 80 % of the excess of the loss over 50, if the loss exceeds 50, and with a maximal reimbursement of 500. Write down a formula for Y in terms of X and draw the graph of this relation.

Question 1.10

For the loss model in Question 1.9, determine a formula for the CDF $F_Y(y)$ and draw its graph.

Question 1.11

For the loss model in Question 1.9, the random variable Y has a mixed distribution. Give a formula for $f_Y(y)$, identifying the discrete and continuous parts.

Question 1.12

Use the results of Question 1.11 to calculate the expected value of the payment per loss Y.

Question 1.13

Duplicate the results of Question 1.12 by viewing Y as a function of X.

Question 1.14

The individual loss amounts against a portfolio are uniformly distributed on $[0, m]$. For 75% of the losses we have $m = 1{,}000$. For the other 25% of losses we have $m = 2{,}000$. Write down a formula for the PDF of a randomly selected loss from this portfolio.

Question 1.15

Using the Double Expectation Theorem, determine the mean and variance for the loss variable in Question 1.14.

Question 1.16

The annual frequency of losses against a policy in a portfolio follows the Poisson probability function:

$$\Pr(N=k|\lambda) = e^{-\lambda}\frac{\lambda^k}{k!} \quad \text{for} \quad k=0,1,2,...$$

Over the portfolio, λ varies uniformly from 0.2 to 1.0. Determine $\Pr(N \geq 2)$ for a randomly selected policy from this portfolio.

Question 1.17

A loss X follows the exponential distribution with mean 50:

$$f_X(x) = \frac{e^{-x/50}}{50} \quad \text{for} \quad x > 0$$

The random variable Y is equal to $1/X$. Calculate the PDF of Y (an **inverse exponential distribution)**.

Question 1.18

You are given:

$$f_1(x) = 0.001 \quad \text{for} \quad 0 < x \leq 1{,}000 \quad , \quad p_1 = 0.80$$

$$f_2(x) = \frac{2 \times 1{,}000^2}{x^3} \quad \text{for} \quad x > 1{,}000 \quad , \quad p_2 = 0.20$$

The distributions are spliced together to form a PDF $f(x)$. Write down a formula for $f(x)$ and compute the mean of the spliced distribution.

Question 1.19

Suppose that X follows the exponential distribution with mean 500:

$$f_X(x) = \frac{e^{-x/500}}{500} \quad \text{for} \quad x > 0$$

If Y is obtained by truncating X below at 100, determine the PDF of Y and its expected value $E\left[X \mid X > 100\right]$.

Question 1.20

Suppose that X follows the exponential distribution with mean 500:

$$f_X(x) = \frac{e^{-x/500}}{500} \quad \text{for} \quad x > 0$$

If Y is obtained by censoring X above at 1,000, determine the PDF of Y (labelling the discrete and continuous parts), and its expected value.

Appendix

Proof

The proof of the DET will assume continuous random variables but the results apply equally well to discrete ones. Note first that we have:

$$E[X \mid Y=y] = \int_{-\infty}^{\infty} x\, f_X(x \mid Y=y)\, dx$$

Now $E[X \mid Y=y]$ is a function of y. So to find its expected value you should multiply $E[X \mid Y=y]$ by the PDF of Y and then integrate to find $E\big[E[X \mid Y=y]\big]$:

$$\begin{aligned} E\big[E[X \mid Y=y]\big] &= \int_{-\infty}^{\infty} \big(E[X \mid Y=y]\big) f_Y(y)\, dy \\ &= \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} x\, f_X(x|Y=y)\, dx\right) f_Y(y)\, dy \\ &= \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} x\, f_X(x|Y=y)\, f_Y(y)\, dx\right) dy \\ &= \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} x\, f_{XY}(x,y)\, dx\right) dy \\ &= E[X] \end{aligned}$$

Since this result holds for every y, we can write $E[X]=E\big[E[X|Y]\big]$.

For the variance formula we proceed as follows. First, note that a simple modification of the argument above (*ie* replacing X by X^2) will result in:

$$E\Big[E\big[X^2|Y\big]\Big] = E\big[X^2\big]$$

Second, since we have

$$\text{var}(X \mid Y) = E\big[X^2|Y\big] - \big(E[X \mid Y]\big)^2,$$

it follows that we have:

$$E\big[X^2|Y\big] = \text{var}(X \mid Y) + \big(E[X \mid Y]\big)^2$$

Finally, we have:

$$\begin{aligned} \text{var}(X) &= E\big[X^2\big] - \big(E[X]\big)^2 \\ &= E\Big[E\big[X^2 \mid Y\big]\Big] - \Big(E\big[E[X \mid Y]\big]\Big)^2 \\ &= E\Big(\text{var}(X \mid Y) + \big(E[X \mid Y]\big)^2\Big) - \Big(E\big[E[X \mid Y]\big]\Big)^2 \\ &= E\big(\text{var}(X \mid Y)\big) + \underbrace{E\Big[\big(E[X \mid Y]\big)^2\Big] - \Big(E\big[E[X \mid Y]\big]\Big)^2}_{\text{var}(E[X|Y])} \end{aligned}$$

□