AGGREGATION OF CORRELATED RISK PORTFOLIOS: MODELS AND ALGORITHMS

SHAUN S. WANG, PH.D.

Abstract

This paper presents a set of tools for modeling and combining correlated risks. Various correlation structures are generated using copula, common mixture, component, and distortion models. These correlation structures are specified in terms of (i) the joint cumulative distribution function or (ii) the joint characteristic function and lend themselves to efficient methods of aggregation by using Monte Carlo simulation or fast Fourier transform.

ACKNOWLEDGEMENT

The author wishes to thank the CAS Committee on Theory of Risk for initiating and supporting this research project. A special thanks goes to Phil Heckman and Glenn Meyers for providing numerous stimulating comments while overseeing the progress of this project. Phil Heckman, Stephen Mildenhall, Don Mango, Louise Francis, and the CAS Editorial Committee, in their thorough review of this paper, have provided many valuable comments which led to significant improvement of an earlier draft report. The author also thanks Ole Hesselager, Julia Wirch, and Ken Seng Tan for comments.

1. INTRODUCTION

A good introduction for this research paper is the original Request For Proposal (RFP) drafted by the CAS Committee on Theory of Risk. In the following paragraph, the original RFP is restated with minor modification.

> Aggregate loss distributions are probability distributions of the total dollar amount of loss under one or

a block of insurance policies. They combine the separate effects of the underlying frequency and severity distributions. In the actuarial literature, a number of methods have been developed for modeling and computing the aggregate loss distributions (see Heckman and Meyers [8], Panjer [21], and Robertson [23]). The main issue underlying this research project is how to combine aggregate loss distributions for separate but correlated classes of business.

Assume a book of business is the union of disjoint classes of business each of which has an aggregate distribution. These distributions may be given in many different ways. Among other ways, they may be specified parametrically, e.g., lognormal or transformed beta with given parameters; they may be given by specifying separate frequency and severity distributions; e.g., negative binomial frequency and Pareto severity with given parameters. The classes of business are *not* independent. For this project, assume that we are given a correlation matrix (or some other easily obtainable measure of dependency) and that the correlation coefficients vary among different pairs of classes. The problem is how to calculate the aggregate loss distribution for the whole book.

In traditional actuarial theory, individual risks are usually assumed to be independent, mainly because the mathematics for correlated risks is less tractable. The CAS recognizes the importance of modeling and combining correlated risks and wishes to enhance the development of tools and models that improve the accuracy of the estimation of aggregate loss distributions for blocks of insurance risks. The modeling of dependent risks has special relevance to the current on-going project of Dynamic Financial Analysis (DFA).

In general, combining correlated loss variables requires knowledge of their joint (multivariate) probability distribution. However, the available data regarding the association between loss variables is often limited to some summary statistics (e.g., correlation matrix). In the special case of a multivariate normal distribution, the covariance matrix and the mean vector, as summary statistics, completely specify the joint distribution. For general loss frequency or severity distributions, specific dependency models have to be used in conjunction with summary statistics. Given fixed marginal distributions and a correlation matrix, one can construct infinitely many joint distributions. Ideally, models for dependency structure should be easy to implement and require relatively few input parameters. As well, the choice of the dependency model and its parameter values should reflect the underlying correlation-generating mechanism.

In developing dependency models, we are aiming at simple implementation by Monte Carlo simulation or by fast Fourier transform. To this end, we will take the following approaches to modeling and combining correlated risks:

Sections 2 to 5 serve as a background before major correlation models are discussed in later parts of this paper. Section 2 reviews some basic concepts for a discrete probability distribution, including probability generating function and fast Fourier transform (FFT). Section 3 reviews the aggregate loss model and the FFT method of calculating aggregate loss distributions. Section 4 introduces some basic concepts and tools for multivariate variables, including the joint cumulative distribution function and the joint probability generating function, which will form the basis of the whole paper. Section 5 reviews some basic measures of dependency, including (Pearson) correlation coefficients, Kendall's tau, and Spearman's rank correlation coefficient.

Sections 6, 7, and 8 investigate various correlation structures by using the concept of copulas (i.e., multivariate uniform distributions) as well as the associated simulation techniques. In

particular, the Cook-Johnson copula and the normal copula lead to efficient simulation techniques.

Sections 9, 10, and 11, with due consideration to the underlying correlation-generating mechanism, present a variety of dependency models by using common mixtures and common shocks. These dependency models allow simple methods of aggregation by Monte Carlo simulation or by fast Fourier transform.

Section 12 presents a multivariate negative binomial model which lends itself to an efficient FFT method of combining the correlated risk portfolios. Section 13 gives an example of this method.

For the reader's convenience, an inventory of commonly used univariate distributions is given in Appendix A, including both discrete and continuous distributions. As a convention, X, Y, and Z represent any random variables (discrete, continuous, or mixed), while N and K represent only discrete variables defined on non-negative integers.

2. PROBABILITY GENERATING FUNCTION AND FFT

This section introduces some basic concepts for discrete probability distributions.

2.1. Discrete Probability Distributions

Let X be a discrete random variable defined on non-negative integers, $0, 1, 2, \ldots$. It may represent

- the *number of claims* arising from a specified block of insurance contracts within a pre-specified time period (such as one year); or
- the *claim amount* from a single claim count, with a pre-specified convenient monetary unit (such as \$1,000).

The random variable X can be fully described by a probability vector

$$\mathbf{f}_X = [f_X(0), f_X(1), f_X(2), \ldots, f_X(R)],$$

or simply

$$\mathbf{f}_X = [f_0, f_1, f_2, \ldots, f_R],$$

with $f_X(i) = f_i = \Pr\{X = i\}$. In this representation, the maximal possible value of X cannot exceed R. When R is finite, X has infinitely many vector representations of the form

$$[f_0, f_1, f_2, \ldots, f_R, 0, 0, \ldots, 0],$$

where a number of zeros are added to the right.

For a discrete variable X with a probability vector $\mathbf{f}_X = [f_0, f_1, f_2, \ldots, f_R]$, the *probability generating function* (p.g.f.) is defined by a symbolic series:

$$P_X(t) = f_0 + f_1 t^1 + f_2 t^2 + f_3 t^3 + \cdots + f_R t^R,$$

which is also the expected value of t^X; i.e., $\mathrm{E}[t^X]$.

EXAMPLE 2.1 If a discrete variable N has the following probabilities

$$\Pr\{N = 0\} = 0.5, \qquad \Pr\{N = 2\} = 0.4, \qquad \Pr\{N = 5\} = 0.1, \tag{2.1}$$

then it can be represented by a vector

$$\mathbf{f}_N = [0.5, 0, 0.4, 0, 0, 0.1, 0, \ldots, 0],$$

and it has a probability generating function

$$P_N(t) = 0.5 + 0.4t^2 + 0.1t^5.$$

EXAMPLE 2.2 If a discrete variable K has the following probabilities

$$\Pr\{K = 1\} = 0.4, \qquad \Pr\{K = 2\} = 0.3, \qquad \Pr\{K = 3\} = 0.3, \tag{2.2}$$

then it can be represented by a vector

$$\mathbf{f}_K = [0, 0.4, 0.3, 0.3, 0, \ldots, 0],$$

and it has a probability generating function

$$P_K(t) = 0.4t + 0.3t^2 + 0.3t^3.$$

2.2. *Fast Fourier Transforms*

First we need to review some basics of complex numbers. Let $\mathrm{i} = \sqrt{-1}$ represent a symbol with the property that $\mathrm{i}^2 = -1$. The complex multiplication is defined as

$$(a + b\mathrm{i})(c + d\mathrm{i}) = (ac - bd) + (ad + bc)\mathrm{i}.$$

An important formula for complex numbers is the Euler formula

$$e^{\mathrm{i}z} = \cos(z) + \mathrm{i}\sin(z).$$

Now we are ready to define the fast Fourier transform. The following description of the FFT method draws on Klugman, Panjer, and Willmot [18] and Brigham [1].

The fast Fourier transform is a one-to-one mapping of n points into n points. For any n-point vector $(f_0, f_1, \ldots, f_{n-1})$, the *fast Fourier transform* is the mapping

$$\mathrm{FFT} : \mathbf{f} = [f_0, f_1, \ldots, f_{n-1}] \mapsto \tilde{\mathbf{f}} = [\tilde{f}_0, \tilde{f}_1, \ldots, \tilde{f}_{n-1}]$$

defined by

$$\tilde{f}_k = \sum_{j=0}^{n-1} f_j \exp\left(\frac{2\pi\mathrm{i}}{n} jk\right), \qquad k = 0, 1, \ldots, n-1. \tag{2.3}$$

This one-to-one mapping has an inverse mapping:

$$f_j = \frac{1}{n} \sum_{k=0}^{n-1} \tilde{f}_k \exp\left(-\frac{2\pi\mathrm{i}}{n} kj\right), \qquad j = 0, 1, \ldots, n-1. \tag{2.4}$$

Note that the inverse fast Fourier transform (IFFT) is almost identical to the FFT except for a sign change and a division by n. In general, the FFT depends on the vector length n.

The fast Fourier transform in Equation 2.3 can also be viewed as a simple matrix multiplication:

$$\tilde{\mathbf{f}} = \mathbf{W}\mathbf{f} = \begin{pmatrix} 1 & 1 & 1 & \cdots & 1 \\ 1 & \omega & \omega^2 & \vdots & \omega^{n-1} \\ 1 & \omega^2 & \omega^4 & \vdots & \omega^{2(n-1)} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & \omega^{n-1} & \omega^{2(n-1)} & \ldots & \omega^{(n-1)^2} \end{pmatrix} \mathbf{f},$$

where $\omega = \exp(2\pi \mathrm{i}/n)$.

The inverse FFT in Equation 2.4 is just $\mathbf{W}^{-1}\tilde{\mathbf{f}}$, where

$$\mathbf{W}^{-1} = \frac{1}{n} \begin{pmatrix} 1 & 1 & 1 & \cdots & 1 \\ 1 & \omega^{-1} & \omega^{-2} & \vdots & \omega^{-(n-1)} \\ 1 & \omega^{-2} & \omega^{-4} & \vdots & \omega^{-2(n-1)} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & \omega^{-(n-1)} & \omega^{-2(n-1)} & \ldots & \omega^{-(n-1)^2} \end{pmatrix}.$$

EXAMPLE 2.3 Reconsider the vector associated with the probability distribution in Equation 2.1. If we use the 5-point vector representation

$$\mathbf{f} = [0.5, 0, 0.4, 0, 0, 0.1],$$

the fast Fourier transform yields

$$\tilde{\mathbf{f}} = [1, 0.35 - 0.2598\mathrm{i}, 0.25 + 0.433\mathrm{i}, \\ 0.8, 0.25 - 0.433\mathrm{i}, 0.35 + 0.2598\mathrm{i}].$$

If we use the 6-point vector representation by (p)adding an additional zero

$$\mathbf{f} = [0.5, 0, 0.4, 0, 0, 0.1, 0],$$

the fast Fourier transform yields a different vector $\tilde{\mathbf{f}}$ as

$$[1, 0.3887 - 0.2925\mathrm{i}, 0.0495 + 0.1302\mathrm{i}, 0.8117 + 0.2345\mathrm{i},$$
$$0.8117 - 0.2345\mathrm{i}, 0.0495 - 0.1302\mathrm{i}, 0.3887 + 0.2925\mathrm{i}].$$

The fast Fourier transform is a "fast" computing algorithm because of the following properties: a fast Fourier transform of length $n = 2^r$ can be rewritten as the sum of two fast Fourier transforms, each of length $n/2 = 2^{r-1}$, the first consisting of the even numbered points and the second the odd numbered points.

$$\begin{aligned}\tilde{f}_k &= \sum_{j=0}^{n-1} f_j \exp\left(\frac{2\pi\mathrm{i}}{n} jk\right) \\ &= \sum_{j=0}^{n/2-1} f_{2j} \exp\left(\frac{2\pi\mathrm{i}}{n} 2jk\right) + \sum_{j=0}^{n/2-1} f_{2j+1} \exp\left(\frac{2\pi\mathrm{i}}{n}(2j+1)k\right) \\ &= \sum_{j=0}^{m-1} f_{2j} \exp\left(\frac{2\pi\mathrm{i}}{m} jk\right) + \exp\left(\frac{2\pi\mathrm{i}}{n} k\right) \sum_{j=0}^{m-1} f_{2j+1} \exp\left(\frac{2\pi\mathrm{i}}{m} jk\right),\end{aligned}$$

where $m = n/2 = 2^{r-1}$. Hence

$$\tilde{f}_k = \tilde{f}_k^a + \exp\left(\frac{2\pi\mathrm{i}}{n} k\right) \tilde{f}_k^b. \tag{2.5}$$

Each of $\tilde{f}_k^a$ and $\tilde{f}_k^b$ can, in turn, be written as the sum of two transforms of length $m/2 = 2^{r-2}$. This can be continued successively.

The successive splitting of transforms into transforms of half the length will result, after r times, in transforms of length 1. Knowing the transform of length 1 will allow one to successively compose the transforms of length 2, 2^2, $2^3, \ldots, 2^r$ by using Equation 2.5.

Based on the above observations, the following comments are in order:

- To fully utilize the FFT speed, it is better to use a probability vector of length $n = 2^r$. This can be easily done by adding a number of zeros to the right.
- Thanks to the fact that many computer packages have already programmed FFT as a built-in function, we don't have to carry out the above steps by ourselves. The main purpose of the above paragraph is to illustrate why FFT is a fast algorithm.

It should be pointed out that many authors define the transform in Equation 2.3 as a *discrete* Fourier transform. The *fast* Fourier transform is simply a method for computing the *discrete* Fourier transform. On the other hand, in some applications such as Microsoft Excel, the term FFT is used to refer to the more general discrete Fourier transform. To simplify the terminology, this paper uses the term FFT for both the transform in Equation 2.3 and the special evaluation technique when $n = 2^r$.

As a theoretical note, the FFT should be viewed as a discretized version of the Fourier transform or characteristic function:

$$\phi(z) = \int_{-\infty}^{\infty} f(x) e^{\mathrm{i}zx}\, dx.$$

The characteristic function maps a continuous probability density function to a complex-valued continuous function, while the FFT maps a vector of n values to a vector of n values of complex numbers. This analog to characteristic functions is crucial to the understanding of various FFT algorithms presented in this paper.

2.3. Convolution

Suppose that N and K are independent discrete random variables defined on non-negative integers. Let $J = N + K$ represent the sum of N and K. The probability distribution of J represents

the *convolution* of the probability distributions of N and K and is defined by

$$\Pr\{J = j\} = \sum_{n=0}^{j} \Pr\{N = n\} \Pr\{K = j - n\}, \qquad j = 0, 1, 2, \ldots .$$

EXAMPLE 2.4 For the random variables defined in Equations 2.1 and 2.2, we have

$$\Pr\{J = 5\} = \Pr\{N + K = 5\} = \sum_{n=0}^{5} \Pr\{N = n\} \Pr\{K = 5 - n\}.$$

Since many of the terms are zero, we have

$$\Pr\{J = 5\} = 0 + 0 + \Pr\{N = 2\}\Pr\{K = 3\} + 0 + 0 + 0 = 0.12.$$

Now let X represent a discrete claim severity distribution defined on non-negative integers. For a fixed number of k claims, the total claim amount has a distribution that can be evaluated through repeated convolutions

$$f_X^{*k}(x) = \sum_{y=0}^{x} f_X^{*(k-1)}(x - y) f_X(y), \qquad x = 1, 2, \ldots, \qquad (2.6)$$

with the convention that

$$f^{*0}(0) = 1.$$

We call f^{*k} the kth fold convolution of f.

2.3.1. Convolution by probability generating function

Note that

$$P_{N+K}(t) = \mathrm{E}[t^{N+K}] = \mathrm{E}[t^N \cdot t^K] = \mathrm{E}[t^N]\mathrm{E}[t^K] = P_N(t) \cdot P_K(t)$$

due to the independence of N and K. In other words, the probability generating function of the sum $N + K$ is the product of $P_N(t)$ and $P_K(t)$.

EXAMPLE 2.5 For the random variables defined in Equations 2.1 and 2.2, in terms of probability generating function we have

$$P_J(t) = P_N(t) \cdot P_K(t) = (0.5 + 0.4t^2 + 0.1t^5)(0.4t + 0.3t^2 + 0.3t^3).$$

After expansion we get

$$P_J(t) = 0.20t + 0.15t^2 + 0.31t^3 + 0.12t^4 + 0.12t^5 + 0.04t^6 + 0.03t^7 + 0.02t^8.$$

The coefficients of t^j give the probability that $J = j$; e.g., $\Pr\{J = 5\} = 0.12$.

2.3.2. *Convolution by FFT*

In terms of a characteristic function we have

$$\phi_{N+K}(t) = \mathrm{E}[e^{\mathrm{i}t(N+K)}] = \mathrm{E}[e^{\mathrm{i}tN} \cdot t^{\mathrm{i}tK}] = \mathrm{E}[t^{\mathrm{i}tN}]\mathrm{E}[t^{\mathrm{i}tK}] = \phi_N(t) \cdot \phi_K(t)$$

due to the independence of N and K. In other words, the characteristic function of the sum $N + K$ is the product of N and K.

Because of this relation in terms of characteristic function, FFT can also be used to perform convolutions. The FFT for the sum of two independent discrete random variables is the product of the FFTs of two individual variables, *provided that enough zeros are added (or padded) to each individual probability vector.* Note that FFT is a one-to-one mapping from n points to n points, which requires that input and output vectors be the same length. On the other hand, a longer vector is generally required for a discrete representation of the sum variable than for each component, since the sum variable will take on larger values with non-zero probability. If there is not enough room in the discrete vector, then the tail probabilities for the sum will wrap around and reappear at the beginning. Therefore, it is crucial

to add enough zeros to the right of each individual probability vector.

2.3.3. FFT Algorithm of convolution

If $\mathbf{f} = [f_0, f_1, \ldots, f_{m-1}]$ and $\mathbf{g} = [g_0, g_1, \ldots, g_{k-1}]$ represent two probability vectors, then the following process can be used to evaluate their convolution:

- Pad zeros to the given vectors $\mathbf{f}$ and $\mathbf{g}$ such that each is of length $n \geq m + k$.
- Apply FFT to each of the vectors: $\tilde{\mathbf{f}} = \text{FFT}(\mathbf{f})$ and $\tilde{\mathbf{g}} = \text{FFT}(\mathbf{g})$.
- Take the product (complex number multiplication), element by element, of the two vectors: $\tilde{\mathbf{h}} = \tilde{\mathbf{f}} \cdot \tilde{\mathbf{g}}$.
- Apply IFFT to $\tilde{\mathbf{h}}$ to recover a probability vector, as the convolution of $\mathbf{f}$ and $\mathbf{g}$.

3. AGGREGATE LOSS MODELS AND THE FFT METHOD

In evaluating insurance losses for a book of business, the frequency/severity approach is the most flexible method, where the estimated mean frequency and mean severity are used to estimate the average aggregate loss. In order to facilitate a dynamic analysis of the underlying risk, the aggregate loss distribution is needed to quantify the inherent variability in the aggregate loss cost. In such situations, in addition to an estimate of the mean frequency and mean severity, probability distributions are needed to describe the possible variations in the number of claims and in the dollar amount of each individual claim. The *aggregate loss distribution* combines the effects of both the claim frequency and claim severity distributions.

This section introduces the basics of aggregate loss models and how FFT can be used to calculate the aggregate loss distribution.

3.1. Claim Frequency Distributions

In modeling the frequency of random processes in many fields of applications, the Poisson distribution is usually the starting point, although the parameter uncertainty regarding the Poisson mean often leads to a negative binomial frequency distribution (see Appendix A.3). Actuaries have found that in most cases the claim frequency can be modeled by the Poisson or negative binomial distributions.

- A Poisson distribution with mean $\lambda > 0$ is defined by a probability function:

$$\Pr\{N = n\} = e^{-\lambda}\frac{\lambda^n}{n!}, \qquad n = 0, 1, 2, \ldots .$$

  The Poisson (λ) distribution has a probability generating function

$$P_N(t) = \mathrm{E}[t^N] = e^{\lambda(t-1)},$$

  with mean and variance both equal to λ; i.e., $\mathrm{E}[N] = \mathrm{Var}[N] = \lambda$.

- A negative binomial distribution, with parameters $\alpha, \beta > 0$, has a probability function:

$$p_n = \Pr\{N = n\} = \frac{\Gamma(\alpha + n)}{\Gamma(\alpha)n!}\left(\frac{1}{1+\beta}\right)^{\alpha}\left(\frac{\beta}{1+\beta}\right)^{n}, \qquad n = 0, 1, 2, \ldots .$$

  It has a probability generating function

$$P_N(t) = [1 - \beta(t-1)]^{-\alpha},$$

  with $\mathrm{E}[N] = \alpha\beta$ and $\mathrm{Var}[N] = \alpha\beta(1+\beta)$. In general, for a negative binomial distribution, the variance exceeds the mean. The variance to mean ratio is

$$\frac{\mathrm{Var}[N]}{\mathrm{E}[N]} = 1 + \beta.$$

Some actuaries consistently use the variance to mean ratio to specify a negative binomial distribution. Heckman and Meyers [8] used the contagion parameter, c, to specify a negative binomial distribution, where

$$\mathrm{Var}[N] = \mathrm{E}[N](1 + c \cdot \mathrm{E}[N]).$$

3.2. Claim Severity Distributions

Models for claim severity are very diverse. In many cases, a theoretical loss distribution is used. A list of the most commonly used theoretical distributions is given in Appendix A, including Pareto, gamma, Weibull and lognormal distributions. Among the commonly used two-parameter distributions, the ordering of heaviness (from most heavy to least heavy) of tails is as follows (see Wang [25]):

Distribution	Ranking
Pareto	1
lognormal	2
exponential inverse Gaussian	3
inverse Gaussian	4
Weibull	5
gamma	6

If a large data set is available, an empirical loss distribution can be used.

Once a severity distribution is selected, in order for fast computer implementation, it is necessary to construct a discrete severity distribution on multiples of a convenient monetary unit h, the *span*. If a theoretical continuous distribution is employed, the following methodology can be used to approximate it by a discrete distribution.

3.2.1. *The rounding method*

Suppose that we are given a continuous distribution with cumulative distribution $F_X(t) = \Pr\{X \le t\}$. Choose a span h as appropriate (such that the number of points are sufficient but not excessive). Let f_j denote the probability placed at jh, $j = 0, 1, 2, \ldots$. Then set

$$\begin{aligned} f_0 &= F_X\left(\frac{h}{2}\right), \\ f_j &= F_X\left(jh + \frac{h}{2}\right) - F_X\left(jh - \frac{h}{2}\right), \qquad j = 1, 2, \ldots . \end{aligned} \tag{3.1}$$

This method splits the probability between $(j + 1)h$ and jh and assigns it to $j + 1$ and j. This, in effect, rounds all amounts to the nearest convenient monetary unit, h, the span of the distribution. For example, the span h can be chosen as every \$1000, \$5000, or \$10,000. As the monetary unit of measurement becomes small, the discrete distribution function needs to approach the true distribution function.

While the main advantage of this rounding method is its simplicity, it has a drawback of not preserving the mean severity of the continuous distribution.

3.2.2. *The matching-mean method*

To avoid the drawback of mismatch of mean severity, one can use a method that forces the matching of the mean.

For a severity distribution with cumulative distribution function F_X, we first evaluate the limited expected values at multiples of h:

$$\mathrm{E}[X; j \cdot h] = \int_0^{j \cdot h} [1 - F_X(u)]\, du, \qquad \text{for} \quad j = 1, 2, \ldots .$$

Then we calculate the probability vector by:

$$f_0 = \Pr\{X = 0 \cdot h\} = 1 - \mathrm{E}[X; h]/h, \tag{3.2}$$

$$\begin{aligned} f_j &= \Pr\{X = j \cdot h\} \\ &= (2\mathrm{E}[X; j \cdot h] - \mathrm{E}[X; (j-1) \cdot h] - \mathrm{E}[X; (j+1) \cdot h])/h, \\ &\qquad\qquad j = 1, 2, \ldots, \end{aligned} \tag{3.3}$$

By doing so, the mean severity of the continuous distribution is preserved in the discrete distribution. One can verify this by taking the sum of f_i, $i = 0, 1, 2, \ldots$.

Recall that by taking the second-order derivative of the limited expected value function we get a probability density function. In the above method, we first obtain a discrete vector of limited expected values; by taking the second-order finite difference, we get a discrete probability function.

3.3. The Aggregation of Frequency and Severity

The aggregate losses are represented as a sum, Z, of a random number, N, of individual payment amounts $(X_1, X_2, \ldots, X_N)$.

The random sum

$$Z = X_1 + X_2 + \cdots + X_N \tag{3.4}$$

has a probability distribution

$$\begin{aligned} f_Z(x) &= \Pr(Z = x) \\ &= \sum_{n=0}^{\infty} \Pr(N = n) \Pr(Z = x \mid N = n) \\ &= \sum_{n=0}^{\infty} \Pr(N = n) f_X^{*n}(x), \end{aligned} \tag{3.5}$$

where $f_X(x) = \Pr(X = x)$ is the common probability distribution of the X_js.

A direct evaluation by Equation 3.5 of the aggregate loss distribution is usually very complicated and time consuming, even with today's fast-speed computers. The next subsection introduces the FFT technique for computing the aggregate loss distribution.

3.3.1. Computing aggregate loss distribution by FFT

In the aggregate loss model in Equation 3.4, we have in terms of characteristic function:

$$\begin{aligned}\phi_Z(t) &= \mathrm{E}[e^{\mathrm{i}t(Z)}] = \mathrm{E}_N[\mathrm{E}[e^{\mathrm{i}t(X_1+\cdots+X_N)} \mid N]] \\ &= \mathrm{E}_N[\phi_X(t)^N] = P_N(\phi_X(t)),\end{aligned}$$

where P_N is the probability generating function of N. This relation in terms of characteristic function suggests the following FFT algorithm for calculating the aggregate loss distribution:

1. Choose $n = 2^r$ for some integer r; n is the number of points desired in the distribution $f_Z(x)$ of aggregate losses. In other words, the aggregate loss distribution has negligible probability outside the range $[0, n]$. This range should be determined before one knows the exact aggregate loss distribution. Knowledge of the mean and standard deviation of the aggregate loss amount should be helpful.

2. Transform the severity probability distribution from a continuous one to a discrete one. The selection of the span h should depend upon the probable range of the severity distribution, as well as the intended application (central range or the extreme right tail). Let $(f_0, f_1, \ldots, f_{m-1})$ represent the discrete claim severity distribution.

   Add zeros to the given severity probability vector so that it is of length n. We denote the padded discrete sever-

ity distribution by

$$\mathbf{f}_X = [f_X(0), f_X(1), \ldots, f_X(n-1)].$$

3. Apply FFT to the severity probability vector: $\tilde{\mathbf{f}}_X = \text{FFT}(\mathbf{f}_X)$.
4. Apply the probability generating function of the frequency, element by element, to the FFT of the severity vector: $\tilde{\mathbf{f}}_Z = P_N(\tilde{\mathbf{f}}_X)$.
5. Apply IFFT to recover the aggregate loss distribution: $\mathbf{f}_Z = \text{IFFT}(\tilde{\mathbf{f}}_X)$.

As a simple example of the above algorithm, let severity be the degenerate distribution \$1 with certainty, and let frequency be negative binomial. Thus the aggregate distribution is the negative binomial. By choosing the number of points, n, the discrete severity distribution is an n-point vector $(0,1,0,\ldots,0)$. The FFT of the severity vector gives a vector of roots of unity. One can check that the FFT algorithm closely reproduces the negative binomial distribution if the number of points used is sufficiently large.

The FFT and IFFT algorithms are available in many computer software packages including Microsoft Excel. This makes the implementation of the FFT method widely accessible.

3.4. Techniques for Combining Multiple Lines of Business

3.4.1. Combining two lines of business by convolution

Suppose that we are combining two lines of business:

- Line 1 has a claim frequency N and a discrete claim severity X.
- Line 2 has a claim frequency K and a discrete claim severity Y.
- Assume that N, X, K, and Y are mutually independent.

- We are interested in the probability distribution of the aggregate losses for the combined portfolio:

$$Z = (X_1 + \cdots + X_N) + (Y_1 + \cdots + Y_K).$$

Under the above assumptions, we have

$$\phi_Z(t) = P_N(\phi_X(t)) \cdot P_K(\phi_Y(t)).$$

This relation in terms of characteristic function suggests the following FFT procedure:

Let $\tilde{\mathbf{g}}$ and $\tilde{\mathbf{h}}$ represent the FFT of the aggregate loss distributions for Line 1 and Line 2, respectively:

$$\tilde{\mathbf{g}} = P_N(\tilde{\mathbf{f}}_X), \qquad \tilde{\mathbf{h}} = P_K(\tilde{\mathbf{f}}_Y).$$

Before applying IFFT to each of $\tilde{\mathbf{g}}$ and $\tilde{\mathbf{h}}$, we take the complex product (element by element) of $\tilde{\mathbf{g}}$ and $\tilde{\mathbf{h}}$. Then apply the IFFT to the product $\tilde{\mathbf{g}} \cdot \tilde{\mathbf{h}}$ to recover the aggregate loss distribution for Line 1 and Line 2 combined:

$$\mathbf{f}_Z = \text{IFFT}(\tilde{\mathbf{g}} \cdot \tilde{\mathbf{h}}).$$

Under this approach, the aggregate claim frequency is the convolution of individual claim frequency distributions. If each individual line has a negative binomial frequency distribution, the aggregate frequency distribution obtained by convolution may no longer be a negative binomial distribution.

3.4.2. The Poisson model

Here is a basic Poisson model for combining different lines of business:

Assume that we are combining k lines of business. For $j = 1,2,\ldots,k$, assume that Line j has a Poisson frequency with mean λ_j and a severity distribution F_j. We assume that losses from different lines of business are *independent*.

In terms of characteristic function we have

$$\phi_Z(t) = \prod_{j=1}^{k} P_{N_j}(\phi_{X_j}(t))$$

$$= \prod_{j=1}^{k} e^{\lambda_j(\phi_{X_j}(t)-1)}$$

$$= e^{\lambda(\phi_X(t)-1)},$$

where $\lambda = \lambda_1 + \cdots + \lambda_k$, and

$$\phi_X(t) = \frac{\lambda_1}{\lambda}\phi_{X_1}(t) + \cdots + \frac{\lambda_k}{\lambda}\phi_{X_k}(t).$$

Therefore, the aggregate losses for the k lines of business combined have a Poisson frequency with mean

$$\lambda = \lambda_1 + \lambda_2 + \cdots + \lambda_k \tag{3.6}$$

and a severity distribution that is a weighted average of each individual severity distribution:

$$F(x) = \frac{\lambda_1}{\lambda}F_1(x) + \frac{\lambda_2}{\lambda}F_2(x) + \cdots + \frac{\lambda_k}{\lambda}F_k(x). \tag{3.7}$$

In summary, under the assumption of mutual independence between lines and a Poisson frequency model for each line, the aggregate loss distribution fork lines can be calculated as if you had a single line, provided that the frequency and severity are adjusted using Equations 3.6 and 3.7.

Next, we must consider the following complications: (i) the presence of parameter risk and (ii) possible correlation between lines. These two factors are often interrelated.

As an alternative to the Poisson model, the negative binomial distribution is commonly used to adjust for parameter uncertainty. Recall that a negative binomial distribution can be ob-

TABLE 1

FREQUENCY/SEVERITY DISTRIBUTIONS

Line	Mean Frequency	Frequency Var/Mean Ratio	Severity Distribution
1	$E(N_1)$	$1 + \beta_1$	F_1
2	$E(N_2)$	$1 + \beta_2$	F_2
⋮	⋮	⋮	⋮
k	$E(N_k)$	$1 + \beta_k$	F_k

tained by assuming a gamma distribution for the unknown Poisson mean (see Appendix A.3).

With the presence of parameter uncertainty, we have to reevaluate the independence assumption between lines. The common parameter uncertainty may have a similar effect (i.e., over- or under-estimate) on our estimates of individual line mean frequencies. In such cases, the individual claim frequencies may be correlated as a result of the common estimation error (due to the same underlying data quality, variations of the underwriting and claim handling practices of an insurer from the industry average, or bias in the trend and development factors used).

3.4.3. Negative binomial model

In general, consider k lines of business with the frequency/severity distributions shown in Table 1.

Regardless of which specific frequency model is used, the following general relationships hold:

- The mean of the aggregate frequency is the sum of each individual line mean frequency:

$$E[N_{agg}] = E[N_1] + E[N_2] + \cdots + E[N_k]. \tag{3.8}$$

- The total variance of the aggregate frequency can be calculated by:

$$\mathrm{Var}[N_{agg}] = \mathrm{Var}\left[\sum_{i=1}^{k} N_i\right] = \sum_{i=1}^{k} \mathrm{Var}[N_i] + 2\sum_{i<j} \mathrm{Cov}[N_i, N_j]. \tag{3.9}$$

A simple and direct approach to incorporating claim count parameter uncertainty is to assume a negative binomial distribution for the aggregate frequency for all lines combined. With this aggregate approach, the negative binomial parameters can be readily estimated from $\mathrm{E}[N_{agg}]$ and $\mathrm{Var}[N_{agg}]$ in Equations 3.8 and 3.9. The severity distribution for all lines combined can be calculated as the weighted average of individual severity distributions:

$$F(x) = \frac{\mathrm{E}[N_1]}{\mathrm{E}[N_{agg}]} F_1(x) + \frac{\mathrm{E}[N_1]}{\mathrm{E}[N_{agg}]} F_2(x) + \cdots + \frac{\mathrm{E}[N_k]}{\mathrm{E}[N_{agg}]} F_k(x). \tag{3.10}$$

Here is the rationale for this approach: Suppose that after applying trend factors and development factors to losses by line of business, we blend all the trended ultimate losses (or in a reinsurance application, losses in an excess layer) from all lines combined. By considering these consolidated losses from all lines of business, the empirical aggregate frequency has a mean as given in Equation 3.8, and the empirical aggregate severity has a severity distribution as in Equation 3.10. The only difference between the aggregate and individual approaches is the following: The individual approach assumes that each line of business has a negative binomial frequency, while the aggregate approach assumes that the aggregate frequency for all lines combined has a negative binomial distribution.

One major advantage of this approach is its simplicity. By simply adjusting the variance to mean ratio in the aggregate negative binomial frequency, one can easily take account of the parameter uncertainty for each line, as well as correlations between

lines. Suppose that we have the following correlation matrix between N_js:

$$\begin{pmatrix} \rho_{11} & \rho_{12} & \cdots & \rho_{1k} \\ \rho_{21} & \rho_{22} & \cdots & \rho_{2k} \\ \vdots & \vdots & \vdots & \vdots \\ \rho_{k1} & \rho_{k2} & \cdots & \rho_{kk} \end{pmatrix},$$

then we can calculate the overall variance to mean ratio using Equation 3.9 and the relation

$$\mathrm{Cov}[N_i, N_j] = \rho_{ij} \sqrt{\mathrm{Var}[N_i]} \sqrt{\mathrm{Var}[N_j]}.$$

The above method may not be theoretically exact if each individual line (instead of the aggregate of all lines) has a negative binomial frequency, as assumed in the Request for Proposal by the CAS Committee on Theory of Risk. Sections 11 and 12 discuss some exact methods for combining individual lines, each having a negative binomial frequency.

3.5. Other Methods for Calculating the Aggregate Loss Distributions

Over the past two decades there have developed a number of methods for calculation of the aggregate loss distribution from given frequency and severity distributions.

1. Panjer's [21] recursive algorithm is easy to explain and implement. In Appendix C we give a brief introduction of this method.
2. The Heckman–Meyers [8] method utilizes direct inversion of characteristic functions.
3. Robertson [23] presented a FFT method using piecewise uniform severity distributions, instead of a discrete severity distribution.

4. The proposed FFT method in this paper uses a discrete severity distribution that has, after padding zeros, $n = 2^r$ points. This is to exploit the fast speed of the FFT algorithm and to facilitate spreadsheet calculations. Another advantage of the FFT method is that it allows a direct extension to multivariate variables, as we will see later in this paper.

5. The recursive method may encounter some numerical problems such as overflow/underflow with a large expected claim count. On the other hand, the Heckman–Meyers method performs well with large claim frequencies. Panjer and Willmot [22] discuss ways of dealing with large frequency problems for the recursive method. For the FFT method, the problem with a large claim count is setting the span small enough to capture features of the severity distribution, but large enough that n times span gives enough room for the aggregate distribution. For divisible frequency distributions like the Poisson and negative binomial, one can get around the problem by building the aggregate distribution in pieces (say a small number of claims at a time) and adding the resulting distributions by convolution.

4. SOME TOOLS FOR MULTIVARIATE DISTRIBUTIONS

4.1. Review of Univariate Case

Let X be a non-negative random variable of discrete, continuous, or mixed type. Let $f_X(x)$ be the probability (density) function of X; i.e.,

$$f_X(x) = \begin{cases} \Pr\{X = x\}, & \text{if } X \text{ is discrete} \\ \dfrac{d}{dx}F_X(x), & \text{if } X \text{ is continuous.} \end{cases}$$

- The *probability generating function* (p.g.f.) of X is defined by

$$P_X(t) = \mathrm{E}[t^X] = \begin{cases} \sum f_X(x)t^x & \text{if } X \text{ is discrete} \\ \int f_X(x)t^x\,dx & \text{if } X \text{ is continuous.} \end{cases}$$

- The *moment generating function* (m.g.f.) of X is defined by

$$M_X(t) = \mathrm{E}[e^{tX}] = P_X(e^t).$$

- The *characteristic function* (ch.f.), also called Fourier transform, is defined by

$$\phi_X(t) = \mathrm{E}[e^{\mathrm{i}tX}] = P_X(e^{\mathrm{i}t}) = M_X(\mathrm{i}t),$$

  where $\mathrm{i} = \sqrt{-1}$ is the imaginary unit.

- It holds that $P_X(1) = M_X(0) = \phi_X(0) = 1$, and

$$\mathrm{E}[X] = \left[\frac{d}{dt}P_X(t)\right]_{t=1} = \left[\frac{d}{dt}M_X(t)\right]_{t=0} = -\mathrm{i}\left[\frac{d}{dt}\phi_X(t)\right]_{t=0}.$$

4.2. Multivariate Framework

For a set of random variables $(X_1,\ldots,X_k)$, let $f_{X_1,\ldots,X_k}$ be their *joint* probability (density) function; i.e.,

$$f_{X_1,\ldots,X_k}(x_1,\ldots,x_k) = \begin{cases} \Pr\{X_1 = x_1,\ldots,X_k = x_k\}, & \text{if the } X_j \text{ are discrete} \\ \dfrac{\partial^k}{\partial x_1\cdots\partial x_k}F_{X_1,\ldots,X_k}(x_1,\ldots,x_k), & \text{if the } X_j \text{ are continuous.} \end{cases}$$

For any subset of $\{X_1,X_2,\ldots,X_k\}$, their (joint) probability distribution is called a *marginal* probability distribution of $f_{X_1,X_2,\ldots,X_k}$. As special cases, f_{X_1} is a univariate marginal distribution of $f_{X_1,X_2,\ldots,X_k}$, and f_{X_1,X_2} is a bivariate marginal distribution of $f_{X_1,X_2,\ldots,X_k}$.

As standard tools for multivariate random variables $(X_1,\ldots,X_k)$, the *joint probability generating function*, *joint moment gen-*

erating function, and *joint characteristic function* are defined as follows (see Johnson et al., [16, pp. 2–12]):

$$P_{X_1,\ldots,X_k}(t_1,\ldots,t_k) = \mathrm{E}[t_1^{X_1}\cdots t_k^{X_k}];$$
$$M_{X_1,\ldots,X_k}(t_1,\ldots,t_k) = \mathrm{E}[e^{t_1X_1+\cdots+t_kX_k}] = P_{X_1,\ldots,X_k}(e^{t_1},\ldots,e^{t_k});$$
$$\phi_{X_1,\ldots,X_k}(t_1,\ldots,t_k) = \mathrm{E}[e^{\mathrm{i}(t_1X_1+\cdots+t_kX_k)}] = P_{X_1,\ldots,X_k}(e^{\mathrm{i}t_1},\ldots,e^{\mathrm{i}t_k}).$$

Note that in terms of the probability (density) function we have

$$P_{X_1,\ldots,X_k}(t_1,\ldots,t_k) =$$
$$\begin{cases} \sum_{(x_1,\ldots,x_k)} f_{X_1,\ldots,X_k}(x_1,\ldots,x_k)t_1^{x_1}\cdots t_k^{x_k}, & \text{discrete case} \\ \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} f_{X_1,\ldots,X_k}(u_1,\ldots,u_k)t_1^{u_1}\cdots t_k^{u_k}\,du_1\cdots du_k, & \text{continuous case.} \end{cases}$$

The joint probability generating function $P_{X_1,\ldots,X_k}$ or the joint characteristic function $\phi_{X_1,\ldots,X_k}$ completely specifies a multivariate distribution. Equivalent results are obtained either in terms of probability generating function or in terms of characteristic function.

- The probability generating function or characteristic function for the univariate marginal distribution F_{X_j} can be obtained by

$$P_{X_j}(t_j) = P_{X_1,\ldots,X_j,\ldots,X_k}(1,\ldots,1,t_j,1,\ldots,1),$$
$$\phi_{X_j}(t_j) = \phi_{X_1,\ldots,X_j,\ldots,X_k}(0,\ldots,0,t_j,0,\ldots,0).$$

- If the variables $X_1,\ldots,X_k$ are mutually independent, then

$$P_{X_1,\ldots,X_k}(t_1,\ldots,t_k) = \prod_{j=1}^{k} P_{X_j}(t_j).$$

- If two sets of variables $\{X_1,\ldots,X_m\}$ and $\{Y_1,\ldots,Y_n\}$ are independent, then

$$P_{X_1,\ldots,X_m,Y_1,\ldots,Y_n}(t_1,\ldots,t_m,s_1,\ldots,s_n) = P_{X_1,\ldots,X_m}(t_1,\ldots,t_m)P_{Y_1,\ldots,Y_n}(s_1,\ldots,s_n).$$

- The covariances can be evaluated by $\mathrm{Cov}[X_i, X_j] = \mathrm{E}[X_i X_j] - \mathrm{E}[X_i]\mathrm{E}[X_j]$ with

$$\mathrm{E}[X_i X_j] = \frac{\partial^2}{\partial t_i \partial t_j} P_{X_1,\ldots,X_m}(1,\ldots,1)$$

$$= -\frac{\partial^2}{\partial t_i \partial t_j} \phi_{X_1,\ldots,X_m}(0,\ldots,0).$$

This can be seen from the expression

$$\frac{\partial^2}{\partial t_i \partial t_j} P_{X_1,\ldots,X_k}(t_1,\ldots,t_k)$$

$$= \sum x_i x_j f_{X_1,\ldots,X_k}(x_1,\ldots,x_k) t_1^{x_1} \cdots t_i^{x_i - 1} \cdots t_j^{x_j - 1} \cdots t_k^{x_k}.$$

- For a discrete multivariate distribution, the joint probability function is

$$f_{X_1,\ldots,X_k}(x_1,\ldots,x_k) = \frac{\partial^{x_1+\cdots+x_k}}{(\partial t_1)^{x_1} \cdots (\partial t_k)^{x_k}} P_{X_1,\ldots,X_k}(0,\ldots,0) \prod_{i=1}^{k} \frac{1}{x_i!}.$$

4.3. Aggregation of Correlated Variables

THEOREM 1 *For any k correlated variables $X_1,\ldots,X_k$ with joint probability generating function $P_{X_1,\ldots,X_k}$ and joint characteristic function $\phi_{X_1,\ldots,X_k}$, the sum $Z = X_1 + \cdots + X_k$ has a probability generating function and a characteristic function:*

$$P_Z(t) = P_{X_1,\ldots,X_k}(t,\ldots,t), \qquad \phi_Z(t) = \phi_{X_1,\ldots,X_k}(t,\ldots,t).$$

Proof $P_Z(t) = \mathrm{E}[t^{X_1+\cdots+X_k}] = \mathrm{E}[t^{X_1} \cdots t^{X_K}] = P_{X_1,\ldots,X_k}(t,\ldots,t).$

If we know the joint characteristic function of the k correlated variables $X_1,\ldots,X_k$, it is straightforward to get the characteristic function for their sum $\phi_Z(t) = \phi_{X_1,\ldots,X_k}(t,\ldots,t)$. Then the probability distribution of Z can be obtained by inverse

Fourier transform. In actual computer implementation, a discrete version (FFT method) can be used. The relation $\phi_{X_1+\cdots+X_k}(t) = \phi_{X_1,\ldots,X_k}(t,\ldots,t)$, along with its associated FFT algorithm, can be used to

- combine correlated risk portfolios if we let X_i represent the aggregate loss distributions for each individual risk portfolio,
- evaluate the total claim number distribution if we let X_i represent the claim frequency for each individual risk portfolio, or
- combine individual claims if we let X_i represent the claim size for each individual risk.

4.4. Aggregation of Risk Portfolios with Correlated Frequencies

Consider the aggregation of two correlated risk portfolios:

$$Z = (X_1 + \cdots + X_N) + (Y_1 + \cdots + Y_K),$$

where N and K are correlated, while the pair (N,K) is independent of the claim sizes X and Y, and the X_is and Y_js are mutually independent. We have

$$\begin{aligned} P_Z(t) &= \mathrm{E}[t^Z] = \mathrm{E}[t^{(X_1+\cdots+X_N)+(Y_1+\cdots+Y_K)}] \\ &= \mathrm{E}_{N,K}\mathrm{E}[t^{(X_1+\cdots+X_n)+(Y_1+\cdots+Y_m)} \mid N = n,\ K = m] \\ &= \mathrm{E}_{N,K}[P_X(t)^N P_Y(t)^K] \\ &= P_{N,K}(P_X(t), P_Y(t)). \end{aligned}$$

In terms of characteristic function we have

$$\phi_Z(t) = P_{N,K}(\phi_X(t), \phi_Y(t)). \tag{4.1}$$

5. MEASURES OF CORRELATION

5.1. Pearson Correlation Coefficients

For random variables X and Y, the *(Pearson) correlation coefficient*, defined by

$$\rho(X,Y) = \frac{\mathrm{Cov}[X,Y]}{\sigma[X]\sigma[Y]},$$

always lies in the range $[-1,1]$. The Pearson correlation coefficient is also called a linear correlation coefficient. Note that $\rho(X,Y) = 1$ if and only if $X = aY + b$ for some constants $a > 0$ and b. If there is no linear relationship between X and Y, the permissible range of $\rho(X,Y)$ is further restricted.

EXAMPLE 5.1 Consider the case that $\log X \sim N(\mu,1)$ and $\log Y \sim N(\mu\sigma,\sigma^2)$. The maximum correlation between X and Y is obtained when the deterministic relation $Y = X^\sigma$ holds. Thus, for random variables with these fixed marginal distributions we have [see Appendix A.4.2]

$$\max\{\rho(X,Y)\} = \frac{\exp(\sigma)-1}{\sqrt{\exp(\sigma^2)-1}\,\sqrt{e-1}}.$$

Observe that

- $\max\{\rho(X,Y)\} = 1$ when $\sigma = 1$ (i.e., $X = Y$),
- $\max\{\rho(X,Y)\}$ decreases to zero as σ increases to ∞, and
- $\max\{\rho(X,Y)\}$ decreases to $1/\sqrt{e-1}$ as σ decreases to 0.

For a set of k random variables $X_1,\ldots,X_k$, the correlation matrix

$$\begin{pmatrix} \rho(X_1,X_1) & \cdots & \rho(X_1,X_k) \\ \vdots & \vdots & \vdots \\ \rho(X_k,X_1) & \cdots & \rho(X_k,X_k) \end{pmatrix}, \qquad -1 \le \rho(X_i,X_j) \le 1,$$

is always positive definite, as it is symmetric and diagonally dominant.

5.2. *Covariance Coefficients*

For non-negative random variables X and Y, we define the *covariance coefficient* as

$$\begin{aligned}\omega(X,Y) &= \frac{\mathrm{Cov}[X,Y]}{\mathrm{E}[X]\mathrm{E}[Y]} = \rho(X,Y)\frac{\sigma[X]}{\mathrm{E}[X]}\frac{\sigma[Y]}{\mathrm{E}[Y]} \\ &= \rho(X,Y)\mathrm{CV}(X)\mathrm{CV}(Y),\end{aligned}$$

where CV refers to the coefficient of variation. Note that the permissible range of $\omega(X,Y)$ depends on the shape of the marginal distributions.

EXAMPLE 5.2 Reconsider the variables X and Y in Example 5.1. It can be shown that

$$\max\{\omega(X,Y)\} = e^{\sigma} - 1.$$

Observe that

- $\max\{\omega(X,Y)\} = e - 1$ when $\sigma = 1$ (i.e., $X = Y$),
- $\max\{\omega(X,Y)\}$ increases to infinity as σ increases to infinity, and
- $\max\{\omega(X,Y)\}$ decreases to zero as σ decreases to zero.

For k non-negative random variables, $X_1,\ldots,X_K$, we define the matrix of covariance coefficients as

$$\begin{pmatrix} \omega(X_1,X_1) & \cdots & \omega(X_1,X_k) \\ \vdots & \vdots & \vdots \\ \omega(X_k,X_1) & \cdots & \omega(X_k,X_k) \end{pmatrix}.$$

One should exercise caution when choosing a parameter value for $\omega(X,Y)$, as its permissible range is sensitive to the marginal

distributions. A practical method for obtaining the maximal positive and negative covariances between risks X and Y is given in Equations 5.1 and 5.2.

5.3. Frechet Bounds, Comonotonicity, and Maximal Correlation

Now consider the bivariate random variables (X,Y). Let

$$F_{X,Y}(x,y) = \Pr\{X \le x, Y \le y\}, \qquad S_{X,Y}(x,y) = \Pr\{X > x,\ Y > y\}$$

be the joint cumulative distribution function and the joint survivor function of (X,Y), respectively. Note that

$$\begin{aligned} F_{X,Y}(x,\infty) &= F_X(x), \\ F_{X,Y}(\infty,y) &= F_Y(y), \qquad \text{for} \quad -\infty < x,y < \infty \\ S_{X,Y}(x,y) &= 1 - F_X(x) - F_Y(y) + F_{X,Y}(x,y) \ne 1 - F_{X,Y}(x,y). \end{aligned}$$

If X and Y are independent, then $F_{X,Y}(x,y) = F_X(x) \cdot F_Y(y)$ and $S_{X,Y}(x,y) = S_X(x) \cdot S_Y(y)$. In general, the joint cumulative distribution function $F(x,y)$ is constrained from above and below.

LEMMA 1 *For any bivariate cumulative distribution function $F_{X,Y}$ with given marginal distributions F_X and F_Y, we have*

$$\max[F_X(x) + F_Y(y) - 1, 0] \le F_{X,Y}(x,y) \le \min[F_X(x), F_Y(y)].$$

Proof The first inequality results from the fact that $S(x,y) \ge 0$, and the second inequality can be proven using $P(A \cap B) \le \min[P(A), P(B)]$.

The upper bound

$$F_u(x,y) = \min[F_X(x), F_Y(y)]$$

and the lower bound

$$F_l(x,y) = \max[F_X(x) + F_Y(y) - 1, 0]$$

are called Frechet bounds.

Closely associated with Frechet bounds is the concept of comonotonicity. The upper Frechet bound is reached if X and Y are comonotonic. The lower Frechet bound is reached if X and $-Y$ are comonotonic.

DEFINITION 1 *Two random variables X and Y are comonotonic if there exists a random variable Z such that*

$$X = u(Z), \qquad Y = v(Z), \qquad \textit{with probability one},$$

where the functions u and v are non-decreasing.

Recall that X and Y are positively perfectly correlated if and only if $Y = aX + b$, $a > 0$. This linear condition is quite restrictive. Comonotonicity is an extension of the concept of perfect correlation to random variables with non-linear relations. Consider the following excess reinsurance arrangement of risk Z: the ceding company retains the first portion of any loss, and the reinsurer pays the excess portion. Putting it mathematically, the payments of the ceding company and the reinsurer will be

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

respectively. Note that X and Y are *not* perfectly correlated since one cannot be written as a function of the other. However, since X and Y are always non-decreasing functions of the original risk Z, they are *comonotonic*. They are bets on the same event, and neither of them is a hedge against the other.

5.4. Comonotonicity and Monte Carlo Simulation

The concept of comonotonicity can also be explained in terms of Monte Carlo simulation by inversion of random uniform numbers.

Assume that X has a cumulative distribution function F_X and a survivor function $S_X(x) = 1 - F_X(x)$. We define F_X^{-1} and S_X^{-1} as

follows:

$$F_X^{-1}(q) = \min\{x : F_X(x) \geq q\}, \qquad 0 < q < 1$$
$$S_X^{-1}(q) = \min\{x : S_X(x) \leq q\}, \qquad 0 < q < 1.$$

Note that F_X^{-1} is non-decreasing, S_X^{-1} is non-increasing, and $S_X^{-1}(q) = F_X^{-1}(1-q)$.

The traditional Monte Carlo simulation method is based on the following result.

LEMMA 2 *For any random variable X and any random variable U which is uniformly distributed on $(0,1)$, X and $F_X^{-1}(U)$ have the same cumulative distribution function.*

Proof $\mathrm{P}\{F_X^{-1}(U) \leq x\} = \mathrm{P}\{U \leq F_X(x)\} = F_X(x)$.

A Monte Carlo simulation of a random variable X can be achieved by first drawing a random uniform number u from $U \sim \mathrm{Uniform}(0,1)$ and then inverting u by $x = F_X^{-1}(u)$.

In order to simulate comonotonic risks X and Y, the same sample of random uniform numbers can be used in an inversion by F_X and F_Y, respectively. By contrast, if X and Y are independent, two independent samples of random uniform numbers have to be used in an inversion by F_X and F_Y, respectively.

For given marginal distributions F_X and F_Y, the maximal possible correlation exists when X and Y are comonotonic. Based on the Monte Carlo method of generating comonotonic risks, we can calculate the maximal possible covariance between two risks with given marginal probability distributions by:

$$\frac{1}{n}\sum_{j=1}^{n} F_X^{-1}\left(\frac{j}{n+1}\right) F_Y^{-1}\left(\frac{j}{n+1}\right) - \left(\frac{1}{n}\sum_{j=1}^{n} F_X^{-1}\left(\frac{j}{n+1}\right)\right)\left(\frac{1}{n}\sum_{j=1}^{n} F_Y^{-1}\left(\frac{j}{n+1}\right)\right) \quad (5.1)$$

for some large number n. The maximal negative correlation exists when X and $-Y$ are comonotonic, in which case an approximation of the covariance can be obtained from

$$\frac{1}{n}\sum_{j=1}^{n} F_X^{-1}\left(\frac{j}{n+1}\right) F_Y^{-1}\left(\frac{n+1-j}{n+1}\right) - \left(\frac{1}{n}\sum_{j=1}^{n} F_X^{-1}\left(\frac{j}{n+1}\right)\right)\left(\frac{1}{n}\sum_{j=1}^{n} F_Y^{-1}\left(\frac{n+1-j}{n+1}\right)\right) \tag{5.2}$$

for some large number n.

As we have seen, the permissible range for the Pearson correlation coefficient can be quite limited and subject to change under a transformation of the random variable. To overcome the shortcomings of the (linear) correlation coefficient, we can use distribution-free measures of correlation such as Kendall's tau and Spearman's rank correlation coefficient.

5.5. Kendall's Tau and Spearman's Rank Correlation Coefficient

Kendall's tau is a nonparametric correlation measure defined as

$$\begin{aligned}\tau &= \tau(X,Y)\\ &= \Pr\{(X_2 - X_1)(Y_2 - Y_1) \geq 0\} - \Pr\{(X_2 - X_1)(Y_2 - Y_1) < 0\},\end{aligned}$$

in which (X_1, Y_1) and (X_2, Y_2) are two independent realizations of a joint distribution.

Another nonparametric correlation measure is Spearman's rank correlation coefficient:

$$\text{RankCorr}(X,Y) = 12\text{E}\left[(F_X(X) - 0.5)(F_Y(Y) - 0.5)\right].$$

Both Kendall's tau and Spearman's rank correlation coefficient satisfy the following properties (see for example, Genest and Mackay [7]):

- $-1 \leq \tau \leq 1$; $-1 \leq \text{RankCorr} \leq 1$,
- if X and Y are comonotonic, then $\tau = 1$ and RankCorr = 1,
- if X and $-Y$ are comonotonic, then $\tau = -1$ and RankCorr = -1,
- if X and Y are independent, then $\tau = 0$ and RankCorr = 0,
- τ is invariant under strictly monotone transforms, that is, if f and g are strictly increasing (or decreasing) functions, then $\tau(f(X), g(Y)) = \tau(X, Y)$ and

$$\text{RankCorr}(f(X), g(Y)) = \text{RankCorr}(X, Y),$$

- if F_X and F_Y are the cumulative distribution functions of two continuous random variables, we have $\tau(F_X(X), F_Y(Y)) = \tau(X, Y)$ and $\text{RankCorr}(F_X(X), F_Y(Y)) = \text{RankCorr}(X, Y)$. Thus, Kendall's tau and rank correlation coefficient are often measured in terms of uniform random variables over $[0,1] \times [0,1]$.

Kendall's tau can be calculated, with due attention to singularity, as

$$\tau(X, Y) = 4 \int_0^1 \int_0^1 F_{X,Y}(x, y) d^2 F_{X,Y}(x, y) - 1.$$

Assume that we have available a random sample of bivariate observations, (X_i, Y_i), $i = 1, \ldots, k$. A non-parametric estimate of Kendall's tau is

$$\hat{\tau}(X, Y) = \frac{2}{k(k-1)} \sum_{i<j} \text{sign}[(X_i - X_j)(Y_i - Y_j)],$$

where sign$[z]$ equals 1, 0, or -1 when z is positive, zero, or negative, respectively.

TABLE 2

A SAMPLE OF INCURRED LOSSES AND ALAE

Claimant #	Amount of Incurred Losses	Amount of ALAE
98001	50	5.0
98002	65	4.0
98003	28	0.0
98004	75	6.5
98005	38	4.5
Average	51.2	4
Std. Dev.	17.15	2.168

EXAMPLE 5.3 Suppose that we have a set of data for incurred losses and allocated loss adjusted expense as shown in Table 2.

The Pearson correlation coefficient can be estimated by:

$$\frac{(50-51.2)(5.0-4.0)+\cdots+(38-51.2)(4.5-4.0)}{5(17.15)(2.168)}=0.78.$$

Kendall's tau can be estimated by

$$\hat{\tau}(X,Y)=\frac{2}{k(k-1)}\sum_{i<j}\text{sign}[(X_i-X_j)(Y_i-Y_j)]=0.6.$$

To calculate the rank correlation coefficient, we first rank each claim by the ordering of losses and ALAE as shown in Table 3.

The Spearman rank correlation coefficient can be calculated as the ordinary Pearson correlation coefficient between the ranks of the losses and ALAE:

$$\frac{(3-3)(4-3)+\cdots+(2-3)(3-3)}{5\sqrt{2}\sqrt{2}}=0.7.$$

The choice between Kendall's tau and the rank correlation coefficient depends on their relative simplicity for the intended application. Some commonly used random number generators

TABLE 3

RANK ORDERING OF LOSSES AND ALAE

Claimant #	Rank of Incurred Losses	Rank of ALAE
98001	3	4
98002	4	2
98003	1	1
98004	5	5
98005	2	3
Median	3	3
Average	3	3
Std. Dev.	$\sqrt{2}$	$\sqrt{2}$

(e.g., Palisade @Risk, which is a Microsoft Excel add-in) have implemented a method from Iman and Conver [12] to induce a given rank correlation structure.

6. THE CONCEPT OF COPULA

Recall that a Monte Carlo simulation of a random variable X can be achieved by first drawing a random uniform number u from $U \sim \text{Uniform}(0,1)$ and then inverting u by $x = F_X^{-1}(u)$. In a similar way, a Monte Carlo simulation of k variables, $(X_1,\ldots,X_k)$, usually starts with k uniform random variables, $(U_1,\ldots,U_k)$. If the variables $(X_1,\ldots,X_k)$ are independent (or correlated), then we need k independent (or correlated) uniform random variables $(U_1,\ldots,U_k)$. For a set of given marginal distributions, the correlation structure of the variables $(X_1,\ldots,X_k)$ is completely determined by the correlation structure of the uniform random variables, $(U_1,\ldots,U_k)$.

DEFINITION 2 *A copula is defined as the joint cumulative distribution function of k uniform random variables*

$$C(u_1,\ldots,u_k) = \Pr\{U_1 \le u_1,\ldots,U_k \le u_k\}.$$

For any set of arbitrary marginal distributions, the formula

$$F_{X_1,\ldots,X_k}(x_1,\ldots,x_k) = C(F_{X_1}(x_1),\ldots,F_{X_k}(x_k)) \tag{6.1}$$

defines a joint cumulative distribution function with marginal cumulative distributions $F_{X_1},\ldots,F_{X_k}$. The formula

$$S_{X_1,\ldots,X_k}(x_1,\ldots,x_k) = C(S_{X_1}(x_1),\ldots,S_{X_k}(x_k)) \tag{6.2}$$

defines a joint survivor function with marginal survivor function $S_{X_1},\ldots,S_{X_k}$.

The multivariate distributions given by Equations 6.1 and 6.2 are usually different, although they both have the same set of Kendall's tau and the same set of rank correlation coefficients.

7. THE COOK–JOHNSON FAMILY OF DISTRIBUTIONS

Let $(U_1,\ldots,U_k)$ be a k-dimensional uniform distribution with support on the hypercube $(0,1)^k$ and having the joint cumulative distribution function

$$F^{(\alpha)}_{U_1,\ldots,U_k}(u_1,\ldots,u_k) = \left\{\sum_{j=1}^{k} u_j^{-1/\alpha} - k + 1\right\}^{-\alpha}, \tag{7.1}$$

where $u_j \in (0,1)$, $j = 1,\ldots,k$, and $\alpha > 0$. This multivariate uniform distribution has a Kendall's tau:

$$\tau(X_i,X_j) = \tau(U_i,U_j) = \frac{1}{1+2\alpha}.$$

On the other hand, for this family of multivariate distributions, there is no simple analytic form for the rank correlation coefficient.

Cook and Johnson [3] studied the family of multivariate uniform distributions given by Equation 7.1. They showed that

$$\lim_{\alpha\to 0} F^{(\alpha)}_{U_1,\ldots,U_k}(u_1,\ldots,u_k) = \min[u_1,\ldots,u_k],$$

and

$$\lim_{\alpha\to\infty} F^{(\alpha)}_{U_1,\ldots,U_k}(u_1,\ldots,u_k) = \prod_{j=1}^{k} u_j.$$

Thus, the correlation approaches its maximum (i.e., comonotonicity) when α decreases to zero, and the correlation approaches zero when α increases to infinity.

Cook and Johnson also gave the following simple simulation algorithm for the multivariate uniform distribution given by Equation 7.1:

STEP 1 Let $Y_1,\ldots,Y_k$ be independent and each have an exponential (1) distribution.

STEP 2 Let Z have a gamma$(\alpha,1)$ distribution.

STEP 3 Then the variables

$$U_j = [1 + Y_j/Z]^{-\alpha}, \qquad j = 1,\ldots,k, \tag{7.2}$$

have a joint cumulative distribution function given by Equation 7.1.

For a set of arbitrary marginal distributions, $F_{X_1},\ldots,F_{X_k}$, we can define a joint cumulative distribution function by

$$F_{X_1,\ldots,X_k}(x_1,\ldots,x_k) = \left\{\sum_{j=1}^{k} F_{X_j}(x_j)^{-1/\alpha} - k + 1\right\}^{-\alpha}. \tag{7.3}$$

Alternatively, we can define a joint survivor function by

$$S_{X_1,\ldots,X_k}(x_1,\ldots,x_k) = \left\{\sum_{j=1}^{k} S_{X_j}(x_j)^{-1/\alpha} - k + 1\right\}^{-\alpha}. \tag{7.4}$$

Note that Kendall's tau for this multivariate distribution is also $1/(1 + 2\alpha)$, which is determined by the underlying copula and is invariant under monotone transforms.

Consider the task of aggregating k risk portfolios, $(X_1,\ldots,X_k)$, where each X_j may represent the aggregate loss amount for the jth risk portfolio. If we assume that $(X_1,\ldots,X_k)$ have a multivariate distribution given by Equation 7.3, a simulation of $(X_1,\ldots,X_k)$ can be easily implemented by:

STEP 4 Invert the $(U_1,\ldots,U_k)$ in Equation 7.2 using $(F_{X_1}^{-1},\ldots, F_{X_k}^{-1})$.

Alternatively, if we assume that $(X_1,\ldots,X_k)$ have a multivariate distribution given by Equation 7.4, a simulation of $(X_1,\ldots,X_k)$ can be easily implemented by:

STEP 4* Invert the $(U_1,\ldots,U_k)$ in Equation 7.2 using $(S_{X_1}^{-1},\ldots, S_{X_k}^{-1})$.

In the multivariate uniform distribution given by Equation 7.1, all correlations are positive. Negative correlations can be accommodated by applying the transforms $U_i^* = 1 - U_i$ to some, but not all, uniform variables in Equation 7.2.

In this dependency model, no restriction is imposed on the marginal distributions, F_{X_j} or S_{X_j}, $j = 1,\ldots,k$. However, the correlation parameters are quite restricted in the sense that the Kendall's taus have to be the same for any pair of risks. To overcome this restriction in the correlation parameters, the *normal copula* permits arbitrary correlation parameters, $\tau_{ij} = \tau(X_i, X_j)$. It is explained in the next section.

8. THE NORMAL COPULA AND MONTE CARLO SIMULATION

In general, the modeling and combining of correlated risks are most straight-forward if the correlated risks have a multivariate normal distribution. In this section, we will use the multivariate normal distribution to construct the normal copula, and then use it to generate multivariate distributions with arbitrary marginal

distributions. The normal copula enjoys much flexibility in the selection of correlation parameters. As well, it lends itself to simple Monte Carlo simulation techniques.

Assume that $(Z_1,\ldots,Z_k)$ have a multivariate normal distribution with standard normal marginal distribution $Z_j \sim N(0,1)$ and a positive definite correlation matrix

$$\mathbf{\Sigma} = \begin{pmatrix} 1 & \rho_{12} & \cdots & \rho_{1k} \\ \rho_{21} & 1 & \cdots & \rho_{2k} \\ \vdots & \vdots & & \vdots \\ \rho_{k1} & \rho_{k2} & \cdots & 1 \end{pmatrix},$$

where $\rho_{ij} = \rho_{ji}$ is the correlation coefficient between Z_i and Z_j. Then $(Z_1,\ldots,Z_k)$ have a joint probability density function:

$$f(z_1,\ldots,z_k) = \frac{1}{\sqrt{(2\pi)^n|\mathbf{\Sigma}|}} \exp\left\{-\tfrac{1}{2}\mathbf{z}'\mathbf{\Sigma}^{-1}\mathbf{z}\right\}, \qquad \mathbf{z} = (z_1,\ldots,z_k). \tag{8.1}$$

From the correlation matrix $\mathbf{\Sigma}$ we can construct a lower triangular matrix

$$\mathbf{B} = \begin{pmatrix} b_{11} & 0 & \cdots & 0 \\ b_{21} & b_{22} & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ b_{k1} & b_{k2} & \cdots & b_{kk} \end{pmatrix},$$

such that $\mathbf{\Sigma} = \mathbf{B}\mathbf{B}'$. In other words, the correlation matrix $\mathbf{\Sigma}$ equals the matrix product of $\mathbf{B}$ and its transpose $\mathbf{B}'$. The elements of the matrix $\mathbf{B}$ can be calculated from the following Choleski's algorithm (see Burden and Faires [2, Section 6.6]; Johnson [17, Section 4.1]):

$$b_{ij} = \frac{\rho_{ij} - \sum_{s=1}^{j-1} b_{is}b_{js}}{\sqrt{1 - \sum_{s=1}^{j-1} b_{js}^2}}, \qquad 1 \le j \le i \le n, \tag{8.2}$$

with the convention that $\sum_{s=1}^{0}(\cdot) = 0$. It is noted that:

- For $i > j$, the denominator of Equation 8.2 equals b_{jj}.
- The elements of **B** should be calculated from top to bottom and from left to right.

The following simulation algorithm can be used to generate multivariate normal variables with a joint probability density function given by Equation 8.1. See Herzog [9], and Fishman [5, pp. 223–224].

STEP 1 Construct the lower triangular matrix $\mathbf{B} = (b_{ij})$ by Equation 8.2.

STEP 2 Generate a column vector of independent standard normal variables $\mathbf{Y} = (Y_1, \ldots, Y_k)'$.

STEP 3 Take the matrix product $\mathbf{Z} = \mathbf{BY}$ of **B** and **Y**. Then $\mathbf{Z} = (Z_1, \ldots, Z_k)'$ has the required joint probability density function given by Equation 8.1.

Let $\Phi(\cdot)$ represent the cumulative distribution function of the standard normal distribution:

$$\Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-t^2/2}\, dt.$$

Then $\Phi(Z_1), \ldots, \Phi(Z_k)$ have a multivariate uniform distribution with Kendall's tau (e.g., Frees and Valdez [6, pp. 25])

$$\tau(\Phi(Z_i), \Phi(Z_j)) = \tau(Z_i, Z_j) = \frac{2}{\pi} \arcsin(\rho_{ij}),$$

and (Spearman) rank correlation coefficient

$$\text{RankCorr}(\Phi(Z_i), \Phi(Z_j)) = \text{RankCorr}(Z_i, Z_j) = \frac{6}{\pi} \arcsin\left(\frac{\rho_{ij}}{2}\right),$$

where $\arcsin(x)$ is an inverse trigonometric function such that $\sin(\arcsin(x)) = x$.

Let's state this result more formally as a theorem due to its importance.

THEOREM 2 *Assume that* $(Z_1,\ldots,Z_k)$ *have a multivariate normal joint probability density function given by Equation* 8.1*, with correlation coefficient* $\rho_{ij} = \rho(Z_i,Z_j)$*. Let* $H(z_1,\ldots,z_k)$ *be their joint cumulative distribution function. Then*

$$C(u_1,\ldots,u_k) = H(\Phi^{-1}(u_1),\ldots,\Phi^{-1}(u_k))$$

defines a multivariate uniform cumulative distribution function—called the normal copula.

For any set of given marginal cumulative distribution functions $F_1,\ldots,F_k$*, the variables*

$$X_1 = F_1^{-1}(\Phi(Z_1)),\ldots,X_k = F_k^{-1}(\Phi(Z_k))$$

have a joint cumulative distribution function

$$F_{X_1,\ldots,X_k}(x_1,\ldots,x_k) = H(\Phi^{-1}(F_1(x_1)),\ldots,\Phi^{-1}(F_k(x_k)))$$

with marginal cumulative distribution functions $F_1,\ldots,F_k$*. The multivariate variables* $(X_1,\ldots,X_k)$ *have Kendall's tau*

$$\tau(X_i,X_j) = \tau(Z_i,Z_j) = \frac{2}{\pi}\arcsin(\rho_{ij})$$

and Spearman's rank correlation coefficients

$$\text{RankCorr}(X_i,X_j) = \text{RankCorr}(Z_i,Z_j) = \frac{6}{\pi}\arcsin\left(\frac{\rho_{ij}}{2}\right).$$

Although the normal copula does not have a simple analytical expression, it lends itself to a very simple Monte Carlo simulation algorithm.

Suppose that we are given a set of correlated risks $(X_1,\ldots,X_k)$ with marginal cumulative distribution functions $F_{X_1},\ldots,F_{X_k}$ and Kendall's tau $\tau_{ij} = \tau(X_i,X_j)$ or rank correlation coefficient

RankCorr(X_i, X_j). If we assume that $(X_1, \ldots, X_k)$ can be described by the normal copula in Theorem 2, then the following Monte Carlo simulation procedure can be used:

STEP 1 Convert the given Kendall's tau or rank correlation coefficient to our usual measure of correlation for multivariate normal variables:

$$\rho_{ij} = \sin\left(\frac{\pi}{2}\tau_{ij}\right) = 2\sin\left(\frac{\pi}{6}\text{RankCorr}(X_i, X_j)\right),$$

and construct the lower triangular matrix $\mathbf{B} = (b_{ij})$ by Equation 8.2.

STEP 2 Generate a column vector of independent standard normal variables $\mathbf{Y} = (Y_1, \ldots, Y_k)'$.

STEP 3 Take the matrix product of $\mathbf{B}$ and $\mathbf{Y}$: $\mathbf{Z} = (Z_1, \ldots, Z_k)' = \mathbf{BY}$.

STEP 4 Set $u_i = \Phi(Z_i)$ for $i = 1, \ldots, k$.

STEP 5 Set $X_i = F_{X_i}^{-1}(u_i)$ for $i = 1, \ldots, k$.

Theorem 2 and the associated simulation algorithm provide a powerful tool for generating correlated variables. The normal copula is very flexible as it allows any (symmetric, positive definite) matrix of rank correlation coefficients (or alternatively, Kendall's tau). The use of this algorithm implicitly assumes that the underlying variables can be described by a normal copula. Of course, there are many correlation structures that differ from a normal copula, for example, the Cook–Johnson distribution in Equation 7.1. In many practical situations, we only have some indication of the correlation parameters without knowing the exact underlying multivariate distribution. In such situations, a normal copula leads to a simple method of simulating the correlated variables.

Appendix B gives an overview of various other families of copulas and the associated Monte Carlo simulation techniques.

9. COMMON MIXTURE MODELS

In many situations, individual risks are correlated since they are subject to the same claim generating mechanism or are influenced by changes in the common underlying economic/legal environment. For instance, in property insurance, risk portfolios in the same geographic location are correlated, where individual claims are contingent on the occurrence and severity of a natural disaster (hurricane, tornado, earthquake, or severe weather condition). In liability insurance, new court rulings or social inflation may set new trends that affect the settlement of all liability claims for one line of business.

One way of modeling situations where the individual risks $\{X_1, X_2, \ldots, X_n\}$ are subject to the same external mechanism is to use a secondary mixing distribution. The uncertainty about the external mechanism is then described by a structure parameter, θ, which can be viewed as a realization of a random variable Θ. The aggregate losses of the risk portfolio can then be seen as a two-stage process: First the external parameter $\Theta = \theta$ is drawn from the distribution function, F_Θ, of Θ. Next, the claim frequency (or severity) of each individual risk X_i $(i = 1, 2, \ldots, n)$ is obtained as a realization from the conditional distribution function, $F_{X_i|\Theta}(x_i \mid \theta)$, of $X_i \mid \Theta$.

9.1. Common Poisson Mixtures

Consider k discrete random variables $N_1, \ldots, N_k$. Assume that there exists a random parameter Θ such that

$$(N_j \mid \Theta = \theta) \sim \text{Poisson}(\theta\lambda_j), \qquad j = 1, \ldots, k,$$

where the variable Θ has a probability density function $\pi(\theta)$ and a moment generating function M_Θ. For any given $\Theta = \theta$, the variables $(N_j \mid \theta)$ are independent and Poisson $(\lambda_j\theta)$ distributed with

a conditional joint probability generating function

$$P_{N_1,\ldots,N_k|\Theta}(t_1,\ldots,t_k \mid \theta) = \mathrm{E}[t_1^{N_1}\cdots t_k^{N_k} \mid \Theta = \theta] = e^{\theta[\lambda_1(t_1-1)+\cdots\lambda_k(t_k-1)]}.$$

However, unconditionally, $N_1,\ldots,N_k$ are correlated as they depend upon the same random parameter Θ. The unconditional joint probability generating function for $N_1,\ldots,N_k$ is

$$\begin{aligned} P_{N_1,\ldots,N_k}(t_1,\ldots,t_k) &= \mathrm{E}_\Theta[\mathrm{E}[t_1^{N_1}\cdots t_k^{N_k} \mid \Theta]] \\ &= \int_0^\infty e^{\theta[\lambda_1(t_1-1)+\cdots+\lambda_k(t_k-1)]}\pi(\theta)\,d\theta \\ &= M_\Theta(\lambda_1(t_1-1)+\cdots+\lambda_k(t_k-1)). \end{aligned}$$

It has marginal probability generating functions $P_{N_j}(t_j) = M_\Theta(\lambda_j(t_j-1))$ with $\mathrm{E}[N_j] = \lambda_j\mathrm{E}[\Theta]$.

Note that

$$\begin{aligned} \mathrm{Cov}[N_i,N_j] &= \mathrm{E}_\Theta\mathrm{Cov}[N_i \mid \Theta, N_j \mid \Theta] + \mathrm{Cov}[\mathrm{E}[N_i \mid \Theta], \mathrm{E}[N_j|\Theta]] \\ &= \mathrm{Cov}[\Theta\lambda_i, \Theta\lambda_j] = \lambda_i\lambda_j\mathrm{Var}[\Theta]. \end{aligned}$$

The covariance coefficient between N_i and N_j $(i \neq j)$ is

$$\omega(N_i,N_j) = \frac{\mathrm{Cov}[N_i,N_j]}{\mathrm{E}[N_i]\mathrm{E}[N_j]} = \frac{\mathrm{Var}[\Theta]}{\{\mathrm{E}[\Theta]\}^2},$$

where ω is the same for all i and j.

EXAMPLE 9.1 If Θ has a gamma$(\alpha,1)$ distribution with moment generating function $M_\Theta(z) = (1-z)^{-\alpha}$, then

$$P_{N_1,\ldots,N_k}(t_1,\ldots,t_k) = [1-\lambda_1(t_1-1)-\cdots-\lambda_k(t_k-1)]^{-\alpha} \tag{9.1}$$

defines a multivariate negative binomial with marginal distributions NB(α,λ_j) and covariance coefficients $\omega(N_i,N_j) = 1/\alpha$.

EXAMPLE 9.2 If Θ has an inverse Gaussian distribution, IG(β,1), with a moment generating function $M_\Theta(z) = e^{1/\beta[1-\sqrt{1-2\beta z}]}$, then

$$P_{N_1,\ldots,N_k}(t_1,\ldots,t_k) = \exp\left\{\frac{1}{\beta} - \frac{1}{\beta}\sqrt{1-2\beta[\lambda_1(t_1-1)+\cdots+\lambda_k(t_k-1)]}\right\}$$

defines a multivariate Poisson inverse Gaussian with marginal distributions P-IG($\beta\lambda_j,\lambda_j$) and covariance coefficients $\omega(N_i,N_j) = \beta$.

Consider combining k risk portfolios. Assume that the frequencies N_j, $j = 1,\ldots,k$, are correlated via a common Poisson-gamma mixture and have a joint probability generating function given by Equation 9.1. If the severities X_j, $j = 1,\ldots,k$, are mutually independent and independent of the frequencies, there is a simple method of combining the aggregate loss distributions. Given $\lambda = \lambda_1 + \cdots + \lambda_k$ and

$$P_X(t) = \frac{\lambda_1}{\lambda}P_{X_1}(t) + \cdots \frac{\lambda_k}{\lambda}P_{X_k}(t),$$

then

$$P_{N_1,\ldots,N_k}(P_{X_1}(t),\ldots,P_{X_k}(t)) = [1-\lambda(P_X(t)-1)]^{-\alpha}.$$

In other words, the total loss amount for the combined risk portfolios has a compound negative binomial distribution with the severity distribution being a weighted average of individual severity distributions. In this case, dependency does not complicate the computation; in fact, it simplifies the calculation. It is simpler than combining independent compound negative binomial distributions.

In this multivariate Poisson-gamma mixture model, the k marginal distributions, negative binomial (α,λ_j), are required to have the same parameter α. This requirement limits its applicability in combining risk portfolios; in many practical cases the frequencies, negative binomial (α_j,λ_j), have different parameter values, α_j. Section 10 and Section 12 overcome this limitation by

extending the Poisson-gamma mixture model to allow arbitrary negative binomial frequencies.

Similar arguments can be made about the Poisson inverse Gaussian distributions.

9.2. *Common Exponential Mixtures*

Consider k continuous random variables $X_1,\ldots,X_k$. Assume that there exists a random parameter Θ such that $(X_j \mid \Theta = \theta)$ is exponentially distributed with parameter $\lambda_j\theta$ and survivor function

$$S_{X_j|\Theta}(t_j \mid \theta) = \Pr\{X_j > t_j \mid \Theta = \theta\} = e^{-\theta\lambda_j t_j}, \qquad j = 1,\ldots,k,$$

where the variable Θ has a probability density function $\pi(\theta)$ and a moment generating function M_Θ.

For any given $\Theta = \theta$, the variables $(X_j \mid \theta)$, $j = 1,\ldots,k$, are conditionally independent and have a conditional joint survivor function

$$\begin{aligned} S_{X_1,\ldots,X_k|\Theta}(t_1,\ldots,t_k \mid \theta) &= \Pr\{X_1 > t_1,\ldots,X_k > t_k \mid \Theta = \theta\} \\ &= e^{-\theta[\lambda_1 t_1 + \cdots + \lambda_k t_k]}. \end{aligned}$$

However, unconditionally, $X_1,\ldots,X_k$ are correlated as they depend upon the same random parameter Θ. The unconditional joint survivor function for $X_1,\ldots,X_k$ is

$$\begin{aligned} S_{X_1,\ldots,X_k}(t_1,\ldots,t_k) &= \int_0^\infty e^{-\theta[\lambda_1 t_1 + \cdots + \lambda_k t_k]}\pi(\theta)\,d\theta \\ &= M_\Theta(-\lambda_1 t_1 - \cdots - \lambda_k t_k). \end{aligned}$$

EXAMPLE 9.3 If Θ has a gamma$(\alpha,1)$ distribution with moment generating function $M_\Theta(z) = (1-z)^{-\alpha}$, this defines a family of multivariate Pareto distributions

$$S_{X_1,\ldots,X_k}(t_1,\ldots,t_k) = [1 + \lambda_1 t_1 + \cdots + \lambda_k t_k]^{-\alpha},$$

with marginal distributions being Pareto$(\alpha, 1/\lambda_j)$.

EXAMPLE 9.4 If Θ has an inverse Gaussian distribution with moment generating function $M_\Theta(z) = e^{1/\beta[1-\sqrt{1-2\beta z}]}$, this defines a family of multivariate exponential inverse Gaussian distributions

$$S_{X_1,\ldots,X_k}(t_1,\ldots,t_k) = \exp\left[\frac{1}{\beta} - \frac{1}{\beta}\sqrt{1 + 2\beta(\lambda_1 t_1 + \cdots + \lambda_k t_k)}\right],$$

with marginal distributions being exponential inverse Gaussian, E-IG$(\beta\lambda_j, \lambda_j)$.

Now we consider the aggregation of k individual claim amounts. Suppose that the k individual claim amounts $X_1,\ldots,X_k$ are identically distributed with $X_i \sim$ Pareto(α,β). But they are correlated by a common exponential-gamma mixture with a joint survivor function

$$S_{X_1,\ldots,X_k}(t_1,\ldots,t_k) = \left[1 + \frac{1}{\beta}(t_1 + \cdots + t_k)\right]^{-\alpha}.$$

Then the sum $X_1 + \cdots + X_k$ has a Pareto$(\alpha, n\beta)$ distribution. This is because, for any given $\Theta = \theta$, $(X_1 + \cdots + X_k \mid \theta) \sim$ exponential(θ/n).

Alternatively, this common exponential mixture model can be obtained by applying the Cook–Johnson copula to k identical marginal survivor functions, Pareto(α,β). In other words, the Cook–Johnson copula can be viewed as an extension of the common exponential mixture model.

10. EXTENDED COMMON POISSON MIXTURE MODELS

The common Poisson mixture model in the previous section has a simple correlation structure and is easy to use. However, it is quite restricted in the sense that it does not permit arbitrary parameter values in the marginal distributions. In this section we extend the common Poisson mixture model so that the marginal distributions may have arbitrary parameter values. This extended model permits simple implementation by Monte Carlo simulation.

Suppose that there exist random variables $(\Theta_1,\ldots,\Theta_k)$ such that for a given set of values $(\Theta_1 = \theta_1,\ldots,\Theta_k = \theta_k)$, the conditional variables $(N_1,\ldots,N_k)$ are independent Poisson(θ_j) variables with

$$P_{N_1,\ldots,N_k|(\Theta_1,\ldots,\Theta_k)}(t_1,\ldots,t_k \mid \theta_1,\ldots,\theta_k)$$
$$= \prod_{j=1}^{k} P_{N_j}(t_j \mid \theta_j) = \prod_{j=1}^{k} e^{\theta_j(t_j-1)},$$

where $M_{\Theta_1,\ldots,\Theta_k}(t_1,\ldots,t_k) = \mathrm{E}_{\Theta_1,\ldots,\Theta_k}[e^{t_1\Theta_1+\cdots+t_k\Theta_k}]$ is the joint moment generating function of $(\Theta_1,\ldots,\Theta_k)$.

The unconditional joint probability generating function is

$$P_{N_1,\ldots,N_k}(t_1,\ldots,t_k) = \mathrm{E}_{(\Theta_1,\ldots,\Theta_k)} P_{N_1,\ldots,N_k}(t_1,\ldots,t_k \mid \Theta_1,\ldots,\Theta_k)$$
$$= M_{\Theta_1,\ldots,\Theta_k}((t_1-1),\ldots,(t_k-1)).$$

By taking the first and second order partial derivatives of this joint probability generating function at $(1,\ldots,1)$, we obtain

$$\mathrm{E}[N_i] = \mathrm{E}[\Theta_i] \qquad \text{and} \qquad \mathrm{Cov}[N_i,N_j] = \mathrm{Cov}[\Theta_i,\Theta_j].$$

We observe a one-to-one correspondence between the correlation structures of the variables $(N_1,\ldots,N_k)$ and the mixing parameters $(\Theta_1,\ldots,\Theta_k)$.

Now consider the case that $\Theta_j \sim \text{gamma}(\alpha_j,\beta_j)$ and thus $N_j \sim \text{NB}(\alpha_j,\beta_j)$, with arbitrary parameter values, $\alpha_j,\beta_j > 0$. We further assume that the variables Θ_j, $j = 1,\ldots,k$, are comonotonic and thus can be simulated by using the same set of uniform random numbers. For $i \neq j$, the covariance $\mathrm{Cov}[\Theta_i,\Theta_j]$ can be numerically calculated by using Equation 5.1. For this dependency model, we have a simple Monte Carlo simulation algorithm:

STEP 1 Generate a uniform number, u, from $U \sim \text{Uniform}(0,1)$.

STEP 2 Let $\theta_j = F^{-1}_{\Theta_j}(u)$, where $\Theta_j \sim \text{gamma}(\alpha_j,\beta_j)$, $j = 1,\ldots,k$.

STEP 3 Simulate $(N_1,\ldots,N_k)$ from k independent Poisson(θ_j) variables, $j = 1,\ldots,k$.

If the α_js are the same, we get the common Poisson mixture model in Example 9.1.

11. COMPONENT MODELS

Consider the aggregation of different lines of business. For a multi-line insurer, the correlation between lines of business may differ from one region to another. Therefore, it may be more appropriate to divide each line into components and model the correlation separately for each component (e.g., by geographic region). There may exist higher correlations between lines in a high catastrophe risk region where the presence of the catastrophe risk may generate a common shock or a common mixture.

Note that many families of frequency and severity distributions are infinitely divisible. A family of distributions is infinitely divisible if any member can be obtained as an independent sum of other members in the same family. Let $X \oplus Y$ represent the sum of two independent random variables and $F_Y \oplus F_Y$ represent the convolution of two probability distributions. We have

- Poisson$(\lambda_1) \oplus$ Poisson(λ_2) = Poisson$(\lambda_1 + \lambda_2)$
- negative binomial: NB$(\alpha_1,\beta) \oplus$ NB(α_2,β) = NB$(\alpha_1 + \alpha_2,\beta)$
- Poisson inverse Gaussian:

$$\text{P-IG}(\beta,\mu_1) \oplus \text{P-IG}(\beta,\mu_2) = \text{P-IG}(\beta,\mu_1 + \mu_2)$$

- gamma$(\alpha_1,\beta) \oplus$ gamma(α_2,β) = gamma$(\alpha_1 + \alpha_2,\beta)$
- inverse Gaussian: IG$(\beta,\mu_1) \oplus$ IG(β,μ_2) = IG$(\beta,\mu_1 + \mu_2)$.

Infinitely divisible distributions are especially useful for dividing risks into independent components. Consider k infinitely

divisible risks $X_j(\alpha_j)$, $j = 1,\ldots,k$, with α_j as the divisible parameter.

Consider a decomposition:

$$\begin{aligned} X_1(\alpha_1) &= X_{11}(\alpha_{11}) \oplus \cdots \oplus X_{1n}(\alpha_{1n}) \\ \vdots \qquad & \qquad \vdots \qquad\qquad \vdots \quad , \qquad \alpha_{js} \geq 0. \quad (11.1) \\ X_k(\alpha_k) &= X_{k1}(\alpha_{k1}) \oplus \cdots \oplus X_{kn}(\alpha_{kn}) \end{aligned}$$

Then we can generate correlation structures component by component:

$$P_{X_1,\ldots,X_k} = \prod_{s=1}^{n} Q_{X_{1s},\ldots,X_{ks}},$$

where the joint probability generating function $Q_{X_{1s},\ldots,X_{ks}}$ for the sth components can be modeled by using a common mixture, a common shock (described below), or by assuming independence, as appropriate. It can be verified that for the component model in Equation 11.1 we have

$$\mathrm{Cov}[X_i, X_j] = \sum_{s=1}^{n} \mathrm{Cov}[X_{is}, X_{js}].$$

11.1. Common Shock Models

Let $X_j = X_{ja} \oplus X_{jb}$, $j = 1,\ldots,k$, be a decomposition into two *independent* components

$$P_{X_1,\ldots,X_k}(t_1,\ldots,t_k) = \mathrm{E}[t_1^{X_{1a}} \cdots t_k^{X_{ka}}]\mathrm{E}[t_1^{X_{1b}} \cdots t_k^{X_{kb}}].$$

If $X_{1a} = \cdots = X_{ka} = X_0$, we obtain

$$P_{X_1,\ldots,X_k}(t_1,\ldots,t_k) = \mathrm{E}[(t_1 \cdots t_k)^{X_0}]\mathrm{E}[t_1^{X_{1b}} \cdots t_k^{X_{kb}}].$$

In particular, if the X_{ib}s are independent, we have $\mathrm{Cov}[X_i, X_j] = \mathrm{Var}[X_0]$. The only source of correlation comes from the common shock variable X_0.

EXAMPLE 11.1 Consider the aggregation of two correlated compound Poisson distributions:

- Portfolio 1. The claim frequency N_1 has a Poisson(λ_1) distribution, and the claim severity X has a probability function $f_1(x)$.
- Portfolio 2. The claim frequency N_2 has a Poisson(λ_2) distribution, and the claim severity Y has a probability function $f_2(y)$.
- Assume that X, Y are independent and both are independent of (N_1, N_2). However, N_1 and N_2 are correlated via a common shock model

$$N_1 = N_0 \oplus N_{1b}, \qquad N_2 = N_0 \oplus N_{2b},$$

where $N_0 \sim \text{Poisson}(\lambda_0)$, $N_{1b} \sim \text{Poisson}(\lambda_1 - \lambda_0)$, and $N_{2b} \sim \text{Poisson}(\lambda_2 - \lambda_0)$.

In this common shock model (N_1, N_2) have a joint probability generating function:

$$\begin{aligned} P_{N_1,N_2}(t_1,t_2) &= \mathrm{E}[t_1^{N_1} t_2^{N_2}] \\ &= \exp[\lambda_1(t_1-1) + \lambda_2(t_2-1) + \lambda_0(t_1-1)(t_2-1)], \end{aligned}$$

with $\text{Cov}[N_1, N_2] = \text{Var}[X_0] = \lambda_0$. It can be shown that the aggregate losses for the combined risk portfolio,

$$S = (X_1 + \cdots + X_{N_1}) + (Y_1 + \cdots + Y_{N_2}),$$

have a compound Poisson($\lambda_1 + \lambda_2 - \lambda_0$) distribution with a severity probability function

$$\begin{aligned} f(z) = \frac{\lambda_1 - \lambda_0}{\lambda_1 + \lambda_2 - \lambda_0} f_1(z) + \frac{\lambda_2 - \lambda_0}{\lambda_1 + \lambda_2 - \lambda_0} f_2(z) \\ + \frac{\lambda_0}{\lambda_1 + \lambda_2 - \lambda_0} f_{1*2}(z), \end{aligned}$$

where f_{1*2} represents the convolution of f_1 and f_2. Thus, existing methods can be applied.

This common shock model can be easily extended to any higher dimension ($k > 2$). For illustrative purposes, we now give an example involving three frequency variables.

EXAMPLE 11.2 The joint probability generating function

$$P_{N_1,N_2,N_3}(t_1,t_2,t_3) = \exp\left\{\sum_{i=1}^{3}\lambda_{ii}(t_i-1)+\sum_{i<j}\lambda_{ij}(t_it_j-1)+\lambda_{123}(t_1t_2t_3-1)\right\} \tag{11.2}$$

defines a multivariate Poisson distribution with marginal distributions

$$N_j \sim \text{Poisson}\left(\lambda_{123}+\sum_{i=1}^{3}\lambda_{ij}\right), \qquad j=1,2,3,$$

and for $i \neq j$, $\text{Cov}[N_i,N_j] = \lambda_{ij} + \lambda_{123}$.

We let

- $K_{ii} \sim \text{Poisson}(\lambda_{ii})$, for $i = 1,2,3$,
- $K_{ij} \sim \text{Poisson}(\lambda_{ij})$, for $1 \leq i < j \leq 3$,
- $K_{ij} = K_{ji}$, for $1 \leq i,j \leq 3$,
- $K_{123} \sim \text{Poisson}(\lambda_{123})$,
- $N_j = K_{1j} \oplus K_{2j} \oplus K_{3j} \oplus K_{123}$, for $j = 1,2,3$.

Then the resulting (N_1,N_2,N_3) have a joint probability generating function given by Equation 11.2. In this model, K_{123} represents the common shock among all three variables (N_1,N_2,N_3). In addition, for $i \neq j$, $K_{ij} = K_{ji}$ represents the extra common shock between N_i and N_j.

Note that we can easily simulate the correlated frequencies, (N_1, N_2, N_3), component by component.

Subject to scale transforms, the common shock multivariate Poisson model can be extended to gamma variables.

EXAMPLE 11.3 Consider two variables $X_1 \sim \text{gamma}(\alpha_1, \beta_1)$ and $X_2 \sim \text{gamma}(\alpha_2, \beta_2)$. Suppose there is a decomposition

$$X_1 = \beta_1(X_0 \oplus X_{1b}), \qquad X_2 = \beta_2(X_0 \oplus X_{2b}),$$

where $X_0 \sim \text{gamma}(\alpha_0, 1)$ with $\alpha_0 \leq \min\{\alpha_1, \alpha_2\}$, $X_{1b} \sim \text{gamma}(\alpha_1 - \alpha_0, 1)$ and $X_{2b} \sim \text{gamma}(\alpha_2 - \alpha_0, 1)$. Then $\text{Cov}[X_1, X_2] = \beta_1\beta_2\text{Var}[X_0] = \alpha_0\beta_1\beta_2$, and

$$X_1 + X_2 = (\beta_1 + \beta_2)X_0 \oplus \beta_1 X_{1b} \oplus \beta_2 X_{2b}.$$

11.2. Peeling Method

Recall that the common Poisson-gamma mixture requires that the marginal distributions $N_j \sim \text{NB}(\alpha, \lambda_j)$ must have the same parameter value α. Now we shall illustrate that, by using the component method, we can construct correlated multivariate negative binomials with arbitrary parameters (α_j, λ_j).

Suppose that we are given k marginal negative binomial distributions:

$$N_1 \sim \text{NB}(\alpha_1, \lambda_1), \ldots, N_k \sim \text{NB}(\alpha_k, \lambda_k).$$

Model 1. Let $\alpha_0 \leq \min\{\alpha_1, \ldots, \alpha_k\}$, and let each N_j $(j = 1, \ldots, k)$ have a decomposition:

$$N_j = N_{ja} \oplus N_{jb}, \qquad N_{ja} \sim \text{NB}(\alpha_0, \lambda_j), \qquad N_{jb} \sim \text{NB}(\alpha_j - \alpha_0, \lambda_j).$$

Note that the N_{ja}s have the same parameter α_0, and thus can be modeled by a common Poisson-gamma mixture

$$P_{N_{1a}, \ldots, N_{ka}}(t_1, \ldots, t_k) = \{1 - \lambda_1(t_1 - 1) - \cdots - \lambda_k(t_k - 1)\}^{-\alpha_0}.$$

If we assume that the N_{jb}s are independent, then $(N_1,\ldots,N_k)$ have a joint probability generating function

$$P_{N_1,\ldots,N_k}(t_1,\ldots,t_k) = \{1-\lambda_1(t_1-1)-\cdots-\lambda_k(t_k-1)\}^{-\alpha_0} \times \prod_{j=1}^{k}\{1-\lambda_j(t_j-1)\}^{\alpha_0-\alpha_j}.$$

Note that

$$\mathrm{Cov}[N_i,N_j] = \alpha_0\lambda_i\lambda_j = \frac{\alpha_0}{\alpha_i\alpha_j}\mathrm{E}[N_i]\mathrm{E}[N_j].$$

Simple methods exist for combining the individual aggregate loss distributions, provided that the severities are mutually independent and independent of $(N_1,\ldots,N_k)$.

Model 2. Assume that the α_j are in an ascending order, $\alpha_1 \le \cdots \le \alpha_k$. The decomposition

$$\mathrm{NB}(\alpha_j,\lambda_j) = \mathrm{NB}(\alpha_1,\lambda_j)\oplus\mathrm{NB}(\alpha_2-\alpha_1,\lambda_j) \oplus\cdots\oplus\mathrm{NB}(\alpha_j-\alpha_{j-1},\lambda_j)$$

can be used in conjunction with common mixture models to generate the following joint probability generating function:

$$P_{N_1,\ldots,N_k}(t_1,\ldots,t_k) = \{1-\lambda_1(t_1-1)-\cdots-\lambda_k(t_k-1)\}^{-\alpha_1} \times\{1-\lambda_2(t_2-1)-\cdots-\lambda_k(t_k-1)\}^{\alpha_1-\alpha_2} \times\cdots\times\{1-\lambda_k(t_k-1)\}^{\alpha_{k-1}-\alpha_k}.$$

It can be verified that the marginal univariate probability generating function is $P_{N_j}(t_j) = [1-\lambda_j(t_j-1)]^{-\alpha_j}$ and the marginal bivariate probability generating function is

$$P_{N_i,N_j}(t_i,t_j) = \{1-\lambda_i(t_i-1)-\lambda_j(t_j-1)\}^{-\alpha_i} \times\{1-\lambda_j(t_j-1)\}^{\alpha_i-\alpha_j}, \qquad i<j,$$

with

$$\mathrm{Cov}[N_i, N_j] = \alpha_i \lambda_i \lambda_j = \frac{1}{\alpha_j} \mathrm{E}[N_i]\mathrm{E}[N_j].$$

11.3. Mixed Correlation Models

Assume that the joint probability generating functions $P_{X_1,\ldots,X_k}$ and $Q_{X_1,\ldots,X_k}$ have the same set of marginal probability generating functions $P_{X_1}, \ldots, P_{X_k}$. Then the mixed joint probability generating function

$$q P_{X_1,\ldots,X_k}(t_1, \ldots, t_k) + (1-q) Q_{X_1,\ldots,X_k}(t_1, \ldots, t_k), \qquad (0 < q < 1),$$

also has marginal probability generating functions $P_{X_1}, \ldots, P_{X_k}$. For this mixed joint probability generating function, we have

$$\mathrm{Cov}[X_i, X_j] = (1-q)\mathrm{Cov}^P[X_i, X_j] + q\,\mathrm{Cov}^Q[X_i, X_j],$$

where Cov^P and Cov^Q represent the covariances implied by the joint probability generating functions P and Q, respectively.

A mixture of joint probability generating functions can be used to represent a set of possible scenarios. For instance, we can let P represent the joint probability generating function under the scenario of major catastrophe occurrence, Q correspond to zero catastrophe occurrence, and q represent the probability of the catastrophe occurrence.

12. THE DISTORTION METHOD

Let $X_1, \ldots, X_k$ be k random variables (discrete, continuous, or multivariate variables) with probability generating functions $P_{X_1}(t_1), \ldots, P_{X_k}(t_k)$, respectively. If the X_js are mutually independent, we have

$$P_{X_1,\ldots,X_k}(t_1, \ldots, t_k) = \prod_{j=1}^{k} P_{X_j}(t_j).$$

Let g be a strictly increasing function over $[0,1]$ with $g(1) = 1$ and whose inverse function is g^{-1}. In a quite loose sense, we

assume that $g \circ P_{X_1,\ldots,X_k}$ specifies a joint probability generating function with marginal probability generating functions $g \circ P_{X_j}$, $(j = 1,\ldots,k)$. By assuming that the distorted joint probability generating function $g \circ P_{X_1,\ldots,X_k}$ has non-correlated marginal probability generating functions, namely,

$$g \circ P_{X_1,\ldots,X_k}(t_1,\ldots,t_k) = \prod_{j=1}^{k} g \circ P_{X_j}(t_j),$$

a correlation structure is introduced to the original joint probability generating function:

$$P_{X_1,\ldots,X_k}(t_1,\ldots,t_k) = g^{-1}\left\{\prod_{j=1}^{k} g \circ P_{X_j}(t_j)\right\}.$$

For mathematical convenience we introduce $h(x) = \ln g(x)$ which is a strictly increasing function over $[0,1]$ with $h(1) = 0$. In terms of h, the above equation can be expressed as

$$P_{X_1,\ldots,X_k}(t_1,\ldots,t_k) = h^{-1}\left\{\sum_{j=1}^{k} h \circ P_{X_j}(t_j)\right\}. \tag{12.1}$$

Note that Equation 12.1 may not define a proper multivariate distribution, as the only constraint on the joint probability (density) function is that it sums to one. It defines a proper multivariate distribution if and only if the joint probability (density) function, $f_{X_1,\ldots,X_k}$, is non-negative everywhere.

Recall that for a discrete distribution,

$$f_{X_1,\ldots,X_k}(x_1,\ldots,x_k) = \frac{\partial^{x_1+\cdots+x_k}}{(\partial t_1)^{x_1}\cdots(\partial t_k)^{x_k}} P_{X_1,\ldots,X_k}(0,\ldots,0)\prod_{i=1}^{k}\frac{1}{x_i!},$$

which can also be derived using multivariate Taylor series expansion. Thus, $P_{X_1,\ldots,X_k}$ defines a proper joint probability distribution if and only if its partial derivatives at $t_1 = \cdots = t_k = 0$ are all non-negative.

THEOREM 3 *Suppose that Equation* 12.1 *defines a joint probability generating function; we have*

$$\mathrm{Cov}[X_i, X_j] = -\left\{\frac{h''(1)}{h'(1)} + 1\right\}\mathrm{E}[X_i]\mathrm{E}[X_j].$$

Proof We take the second order partial derivative, $\partial^2/\partial t_i \partial t_j$, $(i \neq j)$, on both sides of the equation

$$h \circ P_{X_1,\ldots,X_k}(t_1,\ldots,t_k) = \sum_{j=1}^{k} h \circ P_{X_j}(t_j).$$

We obtain zero by taking the second order partial derivative, $\partial^2/\partial t_i \partial t_j$, $(i \neq j)$, on the right-hand side. Thus we should also get zero for the second order partial derivative on the left-hand side:

$$0 = \frac{\partial^2}{\partial t_i \partial t_j}\left\{h \circ P_{X_1,\ldots,X_k}\right\} = \frac{\partial}{\partial t_i}\left\{h'(P_{X_1,\ldots,X_k})\frac{\partial P_{X_1,\ldots,X_k}}{\partial t_j}\right\},$$

which further yields that

$$h''(P_{X_1,\ldots,X_k})\frac{\partial P_{X_1,\ldots,X_k}}{\partial t_i}\frac{\partial P_{X_1,\ldots,X_k}}{\partial t_j} + h'(P_{X_1,\ldots,X_k})\frac{\partial^2 P_{X_1,\ldots,X_k}}{\partial t_i \partial t_j} = 0.$$

Setting the values $t_s = 1$ for $s = 1,\ldots,k$, we get

$$h''(1)\mathrm{E}[X_i]\mathrm{E}[X_j] + h'(1)\mathrm{E}[X_i X_j] = 0.$$

This family of multivariate distributions has a symmetric structure in the sense that ω_{ij} is the same for all $i \neq j$. It would be suitable for combining risks in the same class, where any two individual risks share the same covariance coefficient.

Questions remain as to which distortion function to use and whether the distortion method in Equation 12.1 defines a proper multivariate distribution. In general, the feasibility of the distortion method depends on the marginal distributions.

The next section shows how the distortion method is inherently connected to the common Poisson-mixture models.

12.1. Links with the Common Poisson Mixtures

Reconsider the common Poisson mixture model in Section 9: for any given θ, $(N_j \mid \Theta = \theta)$, $j = 1,\ldots,k$, are conditionally independent Poisson variables with mean $\lambda_j\theta$. If the random parameter Θ has a moment generating function M_Θ, then $(N_1,\ldots,N_k)$ has an unconditional joint probability generating function

$$P_{N_1,\ldots,N_k}(t_1,\ldots,t_k) = M_\Theta(\lambda_1(t_1-1)+\cdots+\lambda_k(t_k-1)),$$

with marginal probability generating function

$$P_{N_j}(t_j) = M_\Theta(\lambda_j(t_j-1)).$$

LEMMA 3 *For a non-negative random variable Θ, the inverse of the moment generating function, M_Θ^{-1}, is well defined over the range* $[0,1]$ *with* $(d/du)M_\Theta^{-1}(u) > 0$, $M_\Theta^{-1}(0) = -\infty$, *and* $M_\Theta^{-1}(1) = 0$.

If we define $h(y) = M_\Theta^{-1}(y)$, then the joint probability generating function for the common Poisson mixture model satisfies

$$P_{N_1,\ldots,N_k}(t_1,\ldots,t_k) = h^{-1}\left\{\sum_{j=1}^{k} h \circ P_{N_j}(t_j)\right\}.$$

EXAMPLE 12.1 If Θ has a gamma$(1/\omega,1)$ distribution with moment generating function $M_\Theta(z) = (1-z)^{-1/\omega}$, then $h(y) = 1 - y^{-\omega}$, and we get the following joint probability generating function:

$$P^{(\omega)}_{N_1,\ldots,N_k}(t_1,\ldots,t_k) = \left\{P_{N_1}(t_1)^{-\omega}+\cdots+P_{N_k}(t_k)^{-\omega}-k+1\right\}^{-1/\omega}, \qquad \omega \neq 0,$$

with

$$\mathrm{Cov}[N_i,N_j] = \omega\,\mathrm{E}[N_i]\mathrm{E}[N_j] \qquad \text{and}$$

$$\lim_{\omega\to 0} P^{(\omega)}_{N_1,\ldots,N_k} = P_{N_1}(t_1)\cdots P_{N_k}(t_k).$$

EXAMPLE 12.2 If Θ has an inverse Gaussian distribution, IG$(\omega, 1)$, with a moment generating function

$$M_\Theta(z) = \exp\left\{\frac{1}{\omega}[1 - \sqrt{1 - 2\omega z}]\right\},$$

then $h(y) = \ln y - \omega/2(\ln y)^2$, and we get the following **joint** probability generating function:

$$P^{(\omega)}_{N_1,\ldots,N_k}(t_1,\ldots,t_k)$$

$$= \exp\left\{\frac{1}{\omega} - \sqrt{\frac{1}{\omega^2} - \sum_{j=1}^{k}\left[\frac{2}{\omega}\ln P_{N_j}(t_j) - (\ln P_{N_j}(t_j))^2\right]}\right\},$$

with

$$\mathrm{Cov}[N_i, N_j] = \omega\,\mathrm{E}[N_i]\mathrm{E}[N_j] \qquad \text{and}$$

$$\lim_{\omega\to 0} P^{(\omega)}_{N_1,\ldots,N_k} = P_{N_1}(t_1)\cdots P_{N_k}(t_k).$$

12.2. A Family of Multivariate Negative Binomial Distributions

As an example of the distortion method, we now discuss a family of multivariate distributions with arbitrary negative binomial marginal distributions, NB(α_j, β_j), $j = 1,\ldots,k$.

THEOREM 4 *The joint probability generating function*

$$P_{N_1,\ldots,N_k}(t_1,\ldots,t_k) = \left\{\sum_{j=1}^{k}[1 - \beta_j(t_j - 1)]^{\alpha_j\omega} - k + 1\right\}^{-1/\omega}, \qquad \omega \neq 0, \tag{12.2}$$

defines a multivariate negative binomial distribution with marginal distributions NB(α_j, β_j) *when either of the following conditions holds:*

- $0 < \omega < \min\{1/\alpha_j,\ j = 1,\ldots,k\}$,

- $\omega < 0$ such that $P_{N_1,\ldots,N_k}(0,\ldots,0) > 0$ and $1/\omega$ is a negative integer.

Proof Equation 12.2 can be rewritten as

$$P_{N_1,\ldots,N_k}(t_1,\ldots,t_k) = Q(t_1,\ldots,t_k)^{-1/\omega},$$

where

$$Q(t_1,\ldots,t_k) = \sum_{j=1}^{k}[1+\beta_j-\beta_j t_j]^{\alpha_j\omega} - k + 1.$$

(i) For $0 < \omega < \min\{1/\alpha_j,\ j = 1,\ldots,k\}$, we have $\alpha_j\omega \le 1$; and the partial derivatives $(\partial^{x_1+\cdots+x_k}/(\partial t_1)^{x_1}\cdots(\partial t_k)^{x_k})P_{N_1,\ldots,N_k}$ are the sum of terms of the following form:

$$aQ(t_1,\ldots,t_k)^{-b}\prod_{j=1}^{k}[1+\beta_j-\beta_j t_j]^{-c_j}, \qquad a,b,c_j \ge 0.$$

Thus, the joint probability function

$$f_{N_1,\ldots,N_k}(x_1,\ldots,x_k) = \frac{\partial^{x_1+\cdots+x_k}}{(\partial t_1)^{x_1}\cdots(\partial t_k)^{x_k}}P_{N_1,\ldots,N_k}(0,\ldots,0)\prod_{i=1}^{k}\frac{1}{x_i!}$$

is always non-negative. Therefore Equation 12.2 does define a proper joint distribution.

(ii) When $\omega < 0$ such that $P_{N_1,\ldots,N_k}(0,\ldots,0) > 0$ and $1/\omega$ is a negative integer, we have

$$P(t_1,\ldots,t_k) = Q(t_1,\ldots,t_k)^n,$$
$$\text{where} \quad n = -1/\omega \text{ is a positive integer,}$$

which can be viewed as the n-fold convolutions of $Q(t_1,\ldots,t_k)$. Note that $[1+\beta_j-\beta_j t_j]^{\alpha_j\omega}$ represents the probability generating function of $\mathrm{NB}(-\alpha_j\omega,\beta_j)$. Thus, $Q(t_1,\ldots,t_k)$ defines a proper multivariate distribution as long as $P_{N_1,\ldots,N_k}(0,\ldots,0) > 0$.

Note that the joint probability generating function in Equation 12.2 requires that ω_{ij} be the same for all i and j, but it allows arbitrary marginal negative binomial distributions, NB(α_j, λ_j). In the special case that all α_j are the same, $\alpha_j = \alpha$, the family of joint distributions in Equation 12.2 returns to the common Poisson-Gamma mixture model with $\omega = 1/\alpha$. This special case corresponds to the usual definition of multivariate negative binomial distributions in Johnson, Kotz and Balakrishnan [16, p. 93]. Thus, Equation 12.2 extends the usual class of multivariate negative binomial distributions.

Remark Consider k individual risk portfolios that are specified by their frequencies and severities: (N_j, X_j), $j = 1,\ldots,k$. Assume that $(N_1,\ldots,N_k)$ has a joint probability generating function as in Equation 12.2, and the only correlation exists between the frequencies. Based on Equation 4.1, the aggregate loss, Z, for the combined risk portfolios has a characteristic function

$$\phi_Z(t) = \left\{ \sum_{j=1}^{k} [1 - \beta_j(\phi_{X_j}(t) - 1)]^{\alpha_j \omega} - k + 1 \right\}^{-1/\omega}, \qquad \omega \neq 0.$$

Thus FFT can be used to evaluate the aggregate loss distribution.

13. AN EXAMPLE OF CORRELATED FREQUENCIES

Consider two correlated risk portfolios with frequency/severity distributions specified as follows:

- Portfolio 1 has a negative binomial frequency with mean = 10 and variance = 20. It has a probability generating function: $P_{N_1}(t) = [1 - (t-1)]^{-10}$. Portfolio 1 has a Pareto($\alpha = 2$, $\beta =$ 50,000) severity subject to a policy limit of \$200,000. Its average severity is \$39,960.
- Portfolio 2 has a negative binomial frequency with mean = 6 and variance = 15. It has a probability generating function: $P_{N_2}(t) = [1 - 2(t-1)]^{-4}$. Portfolio 2 has a Pareto($\alpha = 1.5$, $\beta =$

40,000) severity subject to a policy limit of $300,000. Its average severity is $52,560.

- The two claim frequencies are correlated with a covariance coefficient $\omega_{12} = 0.2$; i.e., $\text{Cov}[N_1, N_2] = 0.2 \cdot \text{E}[N_1] \cdot \text{E}[N_2]$.
- The claim severities X_1 and X_2 for the two risk portfolios are mutually independent; they are also independent of the frequencies (N_1, N_2).

Method I. We approximate the combined frequency N by a negative binomial distribution with

$$\text{E}[N] = \text{E}[N_1] + \text{E}[N_2] = 16,$$

and

$$\text{Var}[N] = \text{Var}[N_1] + \text{Var}[N_2] + 2\text{Cov}[N_1, N_2] = 59.$$

This negative binomial distribution has a probability generating function:

$$P_N(t) = \left[1 - \frac{256}{43}(t-1)\right]^{-43/16}.$$

The combined severity distribution can be calculated by

$$f_X(x) = \frac{\text{E}[N_1]}{\text{E}[N]} f_{X_1}(x) + \frac{\text{E}[N_2]}{\text{E}[N]} f_{X_2}(x),$$

where f_{X_1} and f_{X_2} are the severity distributions for Portfolios 1 and 2, respectively.

Method II. Assume that N_1 and N_2 have a bivariate negative binomial distribution with a joint probability generating function (see Equation 12.2):

$$P_{N_1,N_2}(t_1, t_2) = \{[1 - (t_1 - 1)]^2 + [1 - 2(t_2 - 1)]^{0.8} - 1\}^{-5}.$$

Based on the earlier result in Equation 4.1, the aggregate loss, Z, for the combined risk portfolios has a characteristic function

$$\phi_Z(t) = \{[1 - (\phi_{X_1}(t) - 1)]^2 + [1 - 2(\phi_{X_2}(t) - 1)]^{0.8} - 1\}^{-5}.$$

Thus FFT can be used to evaluate the aggregate loss distribution.

Some details of the calculation steps are as follows:

1. First we approximate the severity distribution by a discrete probability distribution. We choose the number of points for the FFT computation at $4096 = 2^{12}$. This is the maximum number of points for the Microsoft Excel FFT routine. In some other computer languages such as MATLAB, a higher number of points is allowed. We choose a span of $h = \$1{,}000$ and use the "matching-mean" method to approximate each individual severity distribution by a discrete one. For a severity distribution with cumulative distribution function F_X, we first evaluate the limited expected values at multiples of h:

$$\mathrm{E}[X; j \cdot h] = \int_0^{j \cdot h} [1 - F_X(u)]\, du, \qquad \text{for} \quad j = 1, 2, \ldots .$$

Then we apply the following method:

$$\begin{aligned} \Pr\{X = 0 \cdot h\} &= 1 - \mathrm{E}[X; h]/h, \\ \Pr\{X = j \cdot h\} &= (2\mathrm{E}[X; j \cdot h] - \mathrm{E}[X; (j-1) \cdot h] \\ &\qquad - \mathrm{E}[X; (j+1) \cdot h])/h, \qquad j = 1, 2, \ldots . \end{aligned}$$

Note that the severity distribution for the two risk portfolios are subject to policy limits of \$200,000 and \$300,000, respectively. Given that the span was chosen at \$1,000, the maximum severity points with non-zero probabilities are 200 and 300, respectively. *It is critical to pad (i.e., add) enough zeros to the discrete severity vectors so that each severity vector has the same length, 4096 in this case, as the target aggregate loss distribution.* Let $\mathbf{f}_{X_1}$ and $\mathbf{f}_{X_2}$ represent the discrete severity vectors for the two risk portfolios, each of which is of length 4,096.

One should exercise caution in the selection of the span, h, for the discrete severity distributions. Too large a span would affect the accuracy of the discrete distribution. Too small a span may produce some "wrapping" (non-zero probabilities at the high points near 4,096) in the calculated aggregate loss distributions.

2. **Method I:** Let

$$\mathbf{f}_X(j) = \tfrac{10}{16}\mathbf{f}_{X_1}(j) + \tfrac{6}{16}\mathbf{f}_{X_2}(j), \qquad j = 0, 1, \ldots, 4095.$$

Perform FFT on the severity vector $\mathbf{f}_X$. Let $\tilde{\mathbf{f}}_X = \text{FFT}(\mathbf{f}_X)$ represent the resulting vector (of length 4,096). Apply the frequency probability generating function, element by element, to the vector $\tilde{\mathbf{f}}_X$:

$$\tilde{\mathbf{f}}_Z(j) = [1 - \tfrac{256}{43}(\tilde{\mathbf{f}}_X(j) - 1)]^{-43/16}.$$

Finally, perform an inverse FFT on $\tilde{\mathbf{f}}_Z$, and let $\mathbf{f}_Z = \text{IFFT}(\tilde{\mathbf{f}}_Z)$. Note that $\mathbf{f}_Z$ is a probability vector with a span of \$1,000, which approximates the aggregate loss distribution for the combined risk portfolios.

3. **Method II.** Perform FFT on each of the severity vectors, $\mathbf{f}_{X_1}$ and $\mathbf{f}_{X_2}$. Let $\tilde{\mathbf{f}}_{X_1} = \text{FFT}(\mathbf{f}_{X_1})$ and $\tilde{\mathbf{f}}_{X_2} = \text{FFT}(\mathbf{f}_{X_2})$ represent the resulting vectors (each of length 4,096). Apply the bivariate frequency probability generating function:

$$\tilde{\mathbf{f}}_Z(j) = \{[1 - (\tilde{\mathbf{f}}_{X_1}(j) - 1)]^2 + [1 - 2(\tilde{\mathbf{f}}_{X_2}(j) - 1)]^{0.8} - 1\}^{-5}, \qquad j = 0, 1, \ldots, 4095. \tag{13.1}$$

Finally, perform an inverse FFT on $\tilde{\mathbf{f}}_Z$, and let $\mathbf{f}_Z = \text{IFFT}(\tilde{\mathbf{f}}_Z)$. Note that $\mathbf{f}_Z$ is a probability vector with a span of \$1,000, which approximates the aggregate loss distribution for the combined risk portfolios.

4. **Independence Case:** For comparison purposes, we can also calculate the aggregate loss distribution under the assumption of independence between the frequencies. In this case, we repeat Method II except that Equation 13.1 is replaced by the following formula:

$$\tilde{\mathbf{f}}_Z(j) = [1 - (\tilde{\mathbf{f}}_{X_1}(j) - 1)]^{-10} \cdot [1 - 2(\tilde{\mathbf{f}}_{X_2}(j) - 1)]^{-4},$$
$$j = 0, 1, \ldots, 4095.$$

Table 4 lists some values of the calculated aggregate loss distributions.

We can draw two conclusions regarding this specific example:

1. Methods I and II result in two very close aggregate loss distributions.
2. In both methods, correlation has a significant impact on the tail probabilities (quantiles).

14. CONCLUSIONS

This paper has presented a set of tools for modeling and combining correlated risks. A number of correlation structures have been generated using copula, common mixture, component, and distortion models. A good understanding of the claim generating process should be helpful in choosing a model, as well as in selecting correlation parameters. These correlation models are often specified by (i) the joint cumulative distribution function (i.e., a copula) or (ii) the joint characteristic function. The copula construction leads to efficient simulation techniques which can be implemented readily on a spreadsheet. The characteristic function specification leads to simple methods of aggregation by using fast Fourier transforms.

In the high-dimension world of multivariate variables, one may encounter very diverse correlation structures. Regardless of the complexity of the situation, Monte Carlo simulation can al-

TABLE 4

COMPARISON OF VARIOUS METHODS

Loss Amount in Dollars	Method I Single NB	Method II Bivariate NB	Independence Case
x	$\Pr\{Z \le x\}$	$\Pr\{Z \le x\}$	$\Pr\{Z \le x\}$
0	0.00046	0.00032	0.00003
250,000	0.11014	0.11129	0.06888
500,000	0.34756	0.35292	0.30621
750,000	0.59539	0.59897	0.59178
1,000,000	0.77954	0.77937	0.80217
1,250,000	0.89125	0.88894	0.91753
1,500,000	0.95038	0.94777	0.96941
1,750,000	0.97872	0.97672	0.98964
2,000,000	0.99132	0.99006	0.99674
2,250,000	0.99661	0.99590	0.99903
2,500,000	0.99872	0.99836	0.99972
2,750,000	0.99953	0.99936	0.99993
3,000,000	0.99983	0.99976	0.99998
3,250,000	0.99994	0.99991	0.99999
3,500,000	0.99998	0.99997	1.00000
3,750,000	0.99999	0.99999	1.00000
4,000,000	1.00000	1.00000	1.00000
Aggregate Moments	**Method I Single NB**	**Method II Bivariate NB**	**Independence Case**
$\mathrm{E}[Z]$	715,355	715,349	715,361
$\mathrm{CV}[Z]$	0.584	0.593	0.503
$\mathrm{E}[(Z-\mathrm{E}[Z])^3]$	6.948×10^{12}	7.731×10^{12}	3.837×10^{12}

ways be employed in an analysis of the correlation risk. For instance, in some situations, the frequency and severity variables are correlated. With the assistance of Monte Carlo simulation, the common mixture model in Section 9 can be adapted to describe the association between the frequency and severity random variables, if both depend on the same external parameter. This external parameter may be chosen to represent the Richter scale of an earthquake, the velocity of wind speed, or several scenarios of legal climate, etc., depending on the underlying claim environment.

Dependency has always been a fascinating research subject, as well as part of reality. A good understanding of the impact of correlation on the aggregate loss distribution is essential for the dynamic financial analysis of an insurance company. It is hoped that the set of tools developed in this paper will be useful to actuaries in quantifying the aggregate risks of a financial entity. It is also hoped that this research will stimulate more scientific investigations on this subject in the future.

REFERENCES

[1] Brigham, E. O., *The Fast Fourier Transform and Its Applications*, Englewood, NJ, Prentice Hall, 1988.

[2] Burden, R. L., and J. D. Faires, *Numerical Analysis*, 4th edition, Boston, PWS-KENT Publishing Company, 1989.

[3] Cook, R. D., and M. E. Johnson, "A Family of Distributions for Modelling Non-elliptically Symmetric Multivariate Data," *Journal of the Royal Statistical Society*, B43, 1981, pp. 210–218.

[4] Crow, E. L., and K. Shimizu, *Lognormal Distributions: Theory and Applications*, New York, Marcel Dekker, Inc., 1988.

[5] Fishman, G. S., *Monte Carlo: Concepts, Algorithms, and Applications*, New York, Springer-Verlag, 1996.

[6] Frees, E. W., and E. A. Valdez, "Understanding Relationships Using Copulas," *North American Actuarial Journal* 2, 1, 1998, pp. 1–25.

[7] Genest, C., and J. Mackay, "The joy of copulas: bivariate distributions with uniform marginals," *The American Statistician* 40, 4, 1986, pp. 280–283.

[8] Heckman, P. E., and G. G. Meyers, "The Calculation of Aggregate Loss Distributions from Claim Severity and Claim Count Distributions," *PCAS* LXX, 1983, pp. 22–61.

[9] Herzog, T. N., *An Introduction to Stochastic Simulation*, CAS 4B Study Note, 1986.

[10] Hesselager, O., S. S. Wang, and G. E. Willmot, "Exponential and Scale Mixtures and Equilibrium Distribution," *Scandinavian Actuarial Journal* 2, 1997, pp. 125–142.

[11] Hogg, R., and S. Klugman, *Loss Distributions*, New York, John Wiley, 1984.

[12] Iman, R. L., and W. J. Conver, "A Distribution Free Approach to Inducing Rank Correlation Among Input Variables," *Communications in Statistics: Simulation Computation* 11, 3, 1982, pp. 311–334.

[13] Johnson, N., and S. Kotz, *Distributions in Statistics: Continuous Multivariate Distributions*, New York, John Wiley & Sons, Inc., 1972.

[14] Johnson, N., and S. Kotz, "On Some Generalized Farlie-Gumbel-Morgenstern Distributions," *Communications in Statistics* 1977, pp. 415–427.

[15] Johnson, N., and S. Kotz, "On some Generalized Farlie-Gumbel-Morgenstern Distributions–II Regression, Correlation and Further Generalizations," *Communications in Statistics* 1977, pp. 485–496.

[16] Johnson, N., S. Kotz, and N. Balakrishnan, *Discrete Multivariate Distribution*, New York, John Wiley & Sons, Inc., 1997.

[17] Johnson, M. E., *Multivariate Statistical Simulation*, New York, Wiley, 1987.

[18] Klugman, S., H. H. Panjer, and G. E. Willmot, *Loss Models: From Data to Decisions*, New York, John Wiley & Sons, Inc., 1998.

[19] Marshall, A. W., and I. Olkin, "Families of Multivariate Distributions," *Journal of the American Statistical Association* 83, 1988, pp. 834–841.

[20] Oakes, D., "Bivariate Survival Models Induced by Frailties," *Journal of the American Statistical Association* 84, 1989, pp. 487–493.

[21] Panjer, H. H., "Recursive Evaluation of a Family of Compound Distributions," *ASTIN Bulletin* 12, 1981, pp. 22–26.

[22] Panjer, H. H., and G. E. Willmot, "Computational Aspects of Recursive Evaluation of Compound Distributions," *Insurance: Mathematics and Economics* 5, 1986, pp. 113–116.

[23] Robertson, J, "The Computation of Aggregate Loss Distributions," *PCAS* LXXIX, 1992, pp. 57–133.

[24] Vaupel, J. W., K. G. Manton, and E. Stallard, "The Impact of Heterogeneity in Individual Frailty on the Dynamics of Mortality," *Demography* 16, 3, 1979, pp. 439–454.

[25] Wang, S. S., "An Actuarial Index of the Right-Tail Risk," *North American Actuarial Journal* 2, 2, 1998, pp. 88–101.

[26] Willmot, G. E., "The Poisson-Inverse Gaussian as an Alternative to the Negative Binomial," *Scandinavian Actuarial Journal* 1987, pp. 113–127.

APPENDIX A

AN INVENTORY OF UNIVARIATE DISTRIBUTIONS

A.1. Counting Distributions

- The Poisson distribution, Poisson(λ), $\lambda > 0$, is defined by a probability function:

$$p_n = \Pr\{N = n\} = e^{-\lambda}\frac{\lambda^n}{n!}, \qquad n = 0,1,2,\ldots .$$

It has a probability generating function

$$P_N(t) = \mathrm{E}[t^N] = e^{\lambda(t-1)},$$

and $\mathrm{E}[N] = \mathrm{Var}[N] = \lambda$.

- The negative binomial distribution, NB(α,β), $\alpha,\beta > 0$, has a probability function:

$$p_n = \Pr\{N = n\} = \frac{\Gamma(\alpha+n)}{\Gamma(\alpha)n!}\left(\frac{1}{1+\beta}\right)^{\alpha}\left(\frac{\beta}{1+\beta}\right)^{n}, \qquad n = 0,1,2,\ldots .$$

It has a probability generating function

$$P_N(t) = [1-\beta(t-1)]^{-\alpha},$$

with $\mathrm{E}[N] = \alpha\beta$ and $\mathrm{Var}[N] = \alpha\beta(1+\beta)$.

When $\alpha = 1$, the negative binomial distribution NB($1,\beta$) is called the geometric distribution.

- The Poisson inverse Gaussian distribution, P-IG(β,μ), has a probability generating function

$$P_N(t) = \mathrm{E}[t^N] = \exp\left\{-\frac{\mu}{\beta}[\sqrt{1+2\beta(1-t)}-1]\right\}.$$

It can be verified that $\mathrm{E}[N] = \mu$ and $\mathrm{Var}[N] = \mu(1+\beta)$. The probabilities can be calculated via a simple recursion (Willmot, [26]):

$$p_n = \frac{2\beta}{1+2\beta}\left(1 - \frac{3}{2n}\right)p_{n-1} + \frac{\mu^2}{n(n-1)(1+2\beta)}p_{n-2}, \qquad n = 2,3,\ldots,$$

with starting values

$$p_0 = e^{-\mu/\beta[\sqrt{1+2\beta}-1]}, \qquad p_1 = \frac{\mu}{\sqrt{1+2\beta}}p_0.$$

A.2. Continuous Distributions

- The exponential distribution, exponential(λ), is defined by

$$S(x) = 1 - F(x) = e^{-\lambda x}, \qquad x > 0,$$

with $\mathrm{E}[X] = 1/\lambda$ and $\mathrm{Var}[X] = 1/\lambda^2$.

- The gamma distribution, gamma(α,β), $\alpha,\beta > 0$, has a probability density function

$$f(x) = \frac{x^{\alpha-1}e^{-x/\beta}}{\beta^{\alpha}\Gamma(\alpha)}, \qquad x > 0.$$

It has a moment generating function

$$M_X(t) = \mathrm{E}[e^{tX}] = (1-\beta t)^{-\alpha},$$

and $\mathrm{E}[X] = \alpha\beta$, and $\mathrm{Var}[X] = \alpha\beta^2$.

- The Pareto distribution, Pareto(α,β), $\alpha,\beta > 0$, has a survivor function

$$S(x) = 1 - F(x) = \left(\frac{\beta}{x+\beta}\right)^{\alpha} = (1 + x/\beta)^{-\alpha}.$$

The mean $\mathrm{E}[X] = \beta/\alpha - 1$ exists only if $\alpha > 1$.

- The Weibull distribution, Weibull(β, τ), $\beta, \tau > 0$, has a survivor function
$$S(x) = 1 - F(x) = e^{-(x/\beta)^\tau},$$
with $\mathrm{E}[X] = \beta\Gamma(1 + \tau^{-1})$ and $\mathrm{E}[X^2] = \beta^2[\Gamma(1 + 2\tau^{-1})]$.

- The inverse Gaussian distribution, IG(β, μ), has a probability density function
$$f(x) = \mu(2\pi\beta x^3)^{-1/2} \exp\left\{-\frac{(x-\mu)^2}{2\beta x}\right\}, \qquad x > 0.$$
It has a moment generating function
$$M(t) = e^{\mu/\beta[1-\sqrt{1-2\beta t}]},$$
and $\mathrm{E}[X] = \mu$, and $\mathrm{Var}[X] = \mu\beta$.

- The exponential inverse Gaussian distribution, E-IG(β, μ), has a survivor function:
$$S(x) = 1 - F(x) = e^{\mu/\beta\{1-(1+2\beta x)^{1/2}\}}, \qquad x > 0,$$
with moments (Hesselager/Wang/Willmot, [10]):
$$\mathrm{E}[X] = \frac{\beta + \mu}{\mu^2}, \qquad \mathrm{Var}[X] = \frac{5\beta^2 + 4\beta\mu + \mu^2}{\mu^4}.$$

- The lognormal distribution, LN(μ, σ^2), has a probability density function
$$f(x) = \frac{1}{\sqrt{2\pi}\sigma x} \exp\left(-\frac{1}{2}\left[\frac{\log(x) - \mu}{\sigma}\right]^2\right), \qquad x > 0,$$
with
$$\mathrm{E}[X] = \exp[\mu + \sigma^2/2] \qquad \text{and}$$
$$\mathrm{Var}[X] = \exp[2\mu + \sigma^2][\exp(\sigma^2) - 1].$$

A.3. Parameter Uncertainty and Mixture Models

In modeling insurance losses, actuaries are called upon to pick a frequency distribution and a severity distribution based on past claim data and their own judgement. Actuaries are fully aware of the presence of parameter uncertainty in the assumed models. As a way of incorporating parameter uncertainty, mixture models are often employed.

- The most popular frequency distributions are the negative binomial family of distributions. In modeling claim frequency, the negative binomial NB(α, β) can be interpreted as a mixed Poisson model, where the Poisson parameter λ has a gamma(α, β) distribution. This can be seen from the probability generating function

$$\begin{aligned} P_N(t) &= \mathrm{E}[t^N] = \mathrm{E}_\lambda[\mathrm{E}(t^N \mid \lambda)] = \mathrm{E}_\lambda[e^{\lambda(t-1)}] \\ &= M_\lambda(t-1) = \{1 - \beta(t-1)\}^{-\alpha}. \end{aligned}$$

- A popular claim severity distribution is the Pareto distribution which has a thick right tail representing large claims. The Pareto(α, β) distribution can be interpreted as a mixed exponential distribution, where the exponential parameter λ has a gamma$(\alpha, 1/\beta)$ distribution. This can be seen from the survivor function

$$S(x) = \mathrm{E}_\lambda[e^{-\lambda x}] = M_\lambda(-x) = (1 + x/\beta)^{-\alpha} = \left(\frac{\beta}{\beta + x}\right)^{\alpha}.$$

- A more flexible family of claim severity distributions are the Burr distributions (including Pareto as a special case). The Burr(α, β, τ) distributions can be expressed as a Weibull-gamma mixture. This can be seen from the survivor function

$$S(x) = \mathrm{E}_\lambda[e^{-\lambda x^\tau}] = M_\lambda(-x^\tau) = (1 + x^\tau/\beta)^{-\alpha} = \left(\frac{\beta}{\beta + x^\tau}\right)^{\alpha}.$$

  The Burr(α, β, τ) family includes the Pareto(α, β) as a special member when $\tau = 1$.

For $\tau > 1$, the Burr(α, β, τ) distribution has a lighter tail than its Pareto(α, β) counterpart.
For $\tau < 1$, the Burr(α, β, τ) distribution has a thicker tail than its Pareto(α, β) counterpart.

A.4. Lognormal Distributions

A.4.1. Univariate lognormal distributions

The normal distribution, $N(\mu, \sigma^2)$, has a probability density function

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2}\left[\frac{x-\mu}{\sigma}\right]^2\right), \qquad -\infty < x < \infty.$$

It has a moment generating function

$$M_X(t) = \mathrm{E}[e^{tX}] = \exp[\mu t + \tfrac{1}{2}\sigma^2 t^2].$$

If $X \sim N(\mu, \sigma^2)$, then $Y = e^X$ has a lognormal distribution with a probability density function

$$f_Y(y) = \frac{1}{\sqrt{2\pi}\sigma y} \exp\left(-\frac{1}{2}\left[\frac{\log(y)-\mu}{\sigma}\right]^2\right), \qquad y > 0.$$

The moments of Y are

$$\begin{aligned} \mathrm{E}[Y^n] &= \mathrm{E}[\exp(nX)] = M_X(n) \\ &= \exp\left[n\mu + \frac{n^2\sigma^2}{2}\right], \qquad n = 1, 2 \ldots. \end{aligned}$$

Specifically,

$$\begin{aligned} \mathrm{E}[Y] &= \exp\left[\mu + \frac{\sigma^2}{2}\right], \\ \mathrm{Var}[Y] &= \exp[2\mu + \sigma^2][\exp(\sigma^2) - 1], \\ \mathrm{E}[Y - \mathrm{E}[Y]]^3 &= \exp\left[3\mu + \frac{3\sigma^2}{2}\right]\left[\exp(3\sigma^2) - 3\exp(\sigma^2) + 2\right]. \end{aligned}$$

A.4.2. Bivariate lognormal distributions

Let X_1 and X_2 have a bivariate normal distribution with joint probability density function

$$f(x_1,x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \times \exp\left\{-\frac{1}{2(1-\rho)^2}\left[\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2 + \left(\frac{x_2-\mu_2}{\sigma_2}\right)^2 - 2\rho\left(\frac{x_1-\mu_1}{\sigma_1}\right)\left(\frac{x_2-\mu_2}{\sigma_2}\right)\right]\right\}.$$

Then X_1 and X_2 have marginal distributions $N(\mu_1,\sigma_1^2)$ and $N(\mu_2,\sigma_2^2)$, respectively. (X_1,X_2) has a covariance matrix

$$\begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix},$$

where ρ is the correlation coefficient between X_1 and X_2. Note that $\rho = 1$ if and only if $\Pr\{X_1 = aX_2 + b\} = 1$ with $a > 0$.

Now consider the variables $Y_1 = \exp(X_1)$ and $Y_2 = \exp(X_2)$. Note that $\log(Y_1Y_2)$ has a $N(\mu_1+\mu_2, \sigma_1^2+\sigma_2^2+2\rho\sigma_1\sigma_2)$ distribution. We have

$$\begin{aligned}\mathrm{Cov}[Y_1,Y_2] &= \mathrm{E}[Y_1Y_2] - \mathrm{E}[Y_1]\mathrm{E}[Y_2] \\ &= \exp\{(\mu_1+\mu_2) + \tfrac{1}{2}[\sigma_1^2+\sigma_2^2+2\rho\sigma_1\sigma_2]\} \\ &\quad - \exp[\mu_1+\sigma_1^2+\mu_2+\sigma_2^2] \\ &= \exp[\mu_1+\mu_2+\tfrac{1}{2}(\sigma_1^2+\sigma_2^2)]\{\exp(\rho\sigma_1\sigma_2)-1\}.\end{aligned}$$

Therefore, the correlation coefficient of Y_1 and Y_2 is

$$\rho_{Y_1,Y_2} = \frac{\exp(\rho\sigma_1\sigma_2)-1}{\sqrt{\exp(\sigma_1^2)-1}\sqrt{\exp(\sigma_2^2)-1}},$$

where $\rho = \rho_{X_1,X_2}$.

A.4.3. Multivariate lognormal distributions

Consider a vector $(X_1,\dots,X_n)'$ of positive random variables. Assume that $(\log X_1,\dots,\log X_n)'$ has an n-dimensional normal distribution with mean vector and variance-covariance matrix

$$\boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_n \end{pmatrix} \qquad \text{and} \qquad \boldsymbol{\Sigma} = \begin{pmatrix} \sigma_{11} & \cdots & \sigma_{1n} \\ \vdots & \vdots & \vdots \\ \sigma_{n1} & \cdots & \sigma_{nn} \end{pmatrix},$$

respectively.

The distribution of $(X_1,\dots,X_n)'$ is said to be an n-dimensional lognormal distribution with parameters $(\boldsymbol{\mu},\boldsymbol{\Sigma})$ and denoted by $\Lambda_n(\boldsymbol{\mu},\boldsymbol{\Sigma})$. The probability density function of $\mathbf{X} = (X_1,\dots,X_n)'$ having $\Lambda_n(\boldsymbol{\mu},\boldsymbol{\Sigma})$ is (see Crow and Shimizu, [4, Chapter 1]):

$$f(x_1,\dots,x_n) = \begin{cases} \dfrac{1}{\sqrt{(2\pi)^n|\boldsymbol{\Sigma}|}\prod_{i=1}^n x_i} \exp\{-\frac{1}{2}(\log\mathbf{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\log\mathbf{x}-\boldsymbol{\mu})\}, & \mathbf{x}\in(0,\infty)^n \\ 0, & \text{otherwise.} \end{cases}$$

From the moment generating function for the multivariate normal distribution we have

$$\mathrm{E}[X_1^{s_1}\cdots X_n^{s_n}] = \exp(\mathbf{s}'\boldsymbol{\mu} + \tfrac{1}{2}\mathbf{s}'\boldsymbol{\Sigma}\mathbf{s}),$$

where $\mathbf{s} = (s_1,\dots,s_n)'$. Specifically, we have for any $i = 1,2,\dots,n$,

$$\mathrm{E}[X_i^r] = \exp(r\mu_i + \tfrac{1}{2}r^2\sigma_{ii}^2)$$

and for any $i,j = 1,2,\dots,n$,

$$\mathrm{Cov}[X_i,X_j] = \exp\{\mu_i+\mu_j+\tfrac{1}{2}(\sigma_{ii}^2+\sigma_{jj}^2)\}\{\exp(\sigma_{ij})-1\}.$$

A simulation of this multivariate lognormal distribution can be easily achieved by first generating a sample from a multivariate normal distribution and then taking the logarithms.

APPENDIX B

MORE ON COPULAS AND SIMULATION METHODS

This appendix presents greater detail on the construction of copulas and the associated simulation techniques. For simplicity, we confine the discussion to the bivariate case. The discussion here can be readily extended to higher dimensions ($k > 2$).

B.1. Bivariate Copulas

A bivariate *copula* refers to a joint cumulative distribution function $C(u,v) = \Pr\{U \le u; V \le v\}$ with uniform marginals $U, V \sim \text{Uniform}[0,1]$. It links the marginal distributions to their multivariate joint distribution. Let $S_{X,Y}(x,y)$ be a joint survivor function with marginals S_X and S_Y. Then there is a copula C such that

$$S_{X,Y}(x,y) = C(S_X(x), S_Y(y)), \qquad \text{for all} \quad x,y \in (-\infty,\infty).$$

Conversely, given any marginals S_X, S_Y, and a copula C, $S_{X,Y}(x,y) = C(S_X(x), S_Y(y))$ defines a joint survivor function with marginals S_X and S_Y. Furthermore, if S_X and S_Y are continuous, then C is unique.

Note that $S_X(X)$ and $S_Y(Y)$ are uniformly distributed random variables. The association between X and Y can be described by the association between uniform variables $U = S_X(X)$ and $V = S_Y(Y)$. In theory, if one can first generate a sample pair (u_i, v_i) from the bivariate uniform distribution of (U,V), one can simulate a sample pair of (X,Y) by the inverse transforms: $x_i = S_X^{-1}(u_i)$ and $y_i = S_Y^{-1}(v_i)$. Unfortunately, there is no simple way of generating a set of bivariate uniform numbers that works for all copulas. In reality, each type of copula needs a different simulation technique.

Note that

$$F_{X,Y}(x,y) = C(F_X(x), F_Y(y)) \qquad \text{and}$$
$$S_{X,Y}(x,y) = C(S_X(x), S_Y(y))$$

may imply different bivariate distributions. Here we assume that a copula is applied to the survivor functions unless otherwise mentioned.

B.2. Distortion of the Joint Survivor Function

Let $g : [0,1] \to [0,1]$ be an increasing function with $g(0) = 0$ and $g(1) = 1$. If $S_{X,Y}(x,y)$ is a joint survivor function with marginals S_X and S_Y, then $g[S_{X,Y}(x,y)]$ defines another joint survivor function with marginals $g \circ S_X$ and $g \circ S_Y$. If we assume that, after applying a distortion g, $g[S_{X,Y}(x,y)]$ has non-correlated marginals:

$$g[S_{X,Y}(x,y)] = g[S_X(x)]\, g[S_Y(y)],$$

then we have

$$S_{X,Y}(x,y) = g^{-1}(g[S_X(x)] \cdot g[S_Y(y)]), \tag{B.1}$$

which corresponds to the copula

$$C(u,v) = g^{-1}[g(u)g(v)]. \tag{B.2}$$

If we let $h(t) = -\log g(t)$, then Equation B.1 gives the following relation: $S_{X,Y}(x,y) = h^{-1}(h[S_X(x)] + h[S_Y(y)])$, which gives the Archimedean family of copulas (see Genest and Mackay, [7]).

For a bivariate copula C, Kendall's tau is

$$\tau = 4 \int_0^1 \int_0^1 C(u,v)\, dC(u,v) - 1.$$

If a copula $C = g^{-1}(g(u)g(v))$ is defined by a distortion g, then

$$\tau = 1 + 4 \int_0^1 \frac{g(t) \log g(t)}{g'(t)}\, dt.$$

EXAMPLE B.1 The distortion function $g(t) = \exp\{1 - t^{-\alpha}\}$, $\alpha > 0$, corresponds to the Clayton family of copulas with

$$\begin{aligned} C_\alpha(u,v) &= \{u^{-\alpha} + v^{-\alpha} - 1\}^{-1/\alpha}, \\ C_\infty(u,v) &= \lim_{\alpha\to\infty} C_\alpha(u,v) = \min[u,v], \\ C_0(u,v) &= \lim_{\alpha\to 0+} C_\alpha(u,v) = uv. \end{aligned} \tag{B.3}$$

Thus, C_∞ and C_0 are the copulas for the Frechet upper bounds and the independent case, respectively.

EXAMPLE B.2 The distortion function $g(t) = \exp\{-(-\log t)^\alpha\}$, $\alpha \geq 1$, corresponds to the Hougaard family of copulas with

$$\begin{aligned} C_\alpha(u,v) &= \exp\{-[(-\log u)^\alpha + (-\log v)^\alpha]^{1/\alpha}\}, \\ C_\infty(u,v) &= \lim_{\alpha\to\infty} C_\alpha(u,v) = \min[u,v], \\ C_1(u,v) &= uv. \end{aligned} \tag{B.4}$$

Thus, C_∞ and C_1 are the copulas for the Frechet upper bounds and the independent case, respectively.

EXAMPLE B.3 The distortion function $g(t) = \alpha^t - 1/\alpha - 1$, $\alpha > 0$, corresponds to the Frank family of copulas with

$$\begin{aligned} C_\alpha(u,v) &= [\log\alpha]^{-1} \log\left\{1 + \frac{(\alpha^u - 1)(\alpha^v - 1)}{\alpha - 1}\right\}, \qquad 0 < \alpha < \infty, \\ C_\infty(u,v) &= \lim_{\alpha\to\infty} C_\alpha(u,v) = \max[u + v - 1, 0], \\ C_0(u,v) &= \lim_{\alpha\to 0+} C_\alpha(u,v) = \min[u,v], \\ C_1(u,v) &= \lim_{\alpha\to 1} C_\alpha(u,v) = uv. \end{aligned} \tag{B.5}$$

Thus, C_∞, C_0, and C_1 are the copulas for the Frechet lower and upper bounds and the independent case, respectively.

B.3. Common Frailty Models

Vaupel, Manton, and Stallard [24] introduced the concept of *frailty* in their discussion of a heterogeneous population. Each individual in the population is associated with a frailty, r. The frailty varies across individuals and thus is modeled as a random variable R with cumulative distribution function $F_R(r)$. The conditional survival function of lifetime T, given r, is

$$\Pr\{T > t \mid R = r\} = B(t)^r,$$

in which $B(t)$ is the base line survivor function (for a standard individual with $r = 1$). The unconditional survivor function for a heterogeneous population is

$$\Pr\{T > t\} = \int_0^\infty B(t)^r \, dF_R(r) = M_R(\log B(t)),$$

where M_R is the moment generating function of R.

Oakes [20] uses a bivariate frailty model to describe associations between two random variables as follows. Assume that X and Y both can be modeled by frailty models

$$S_X(x) = \int_0^\infty B_1(x)^r \, dF_R(r) = M_R(\log B_1(x)) \qquad \text{and}$$

$$S_Y(y) = \int_0^\infty B_2(y)^r \, dF_R(r) = M_R(\log B_2(y)),$$

respectively, in which B_1 and B_2 are the base line survivor functions. Assume that X and Y are conditionally independent, given frailty $R = r$. However, X and Y are associated as they depend on the common frailty variable R. The bivariate frailty model yields the following joint survivor function

$$\begin{aligned} S_{X,Y}(x,y) = \Pr\{X > x, Y > y\} &= \int_0^\infty [B_1(x) \cdot B_2(y)]^r \, dF_R(r) \\ &= M_R(\log[B_1(x) \cdot B_2(y)]). \end{aligned}$$

For $g(u) = \exp[M_R^{-1}(u)]$, we have

$$g[S_{X,Y}(x,y)] = B_1(x) \cdot B_2(y) = g[S_X(x)] \cdot g[S_Y(y)].$$

In other words, a bivariate frailty model yields a joint distribution that can also be obtained by using the distortion function g.

EXAMPLE B.4 Assume that the frailty R has a gamma distribution with $M_R(z) = (1/1-z)^{1/\alpha},\ \alpha > 0$. Then $M_R^{-1}(u) = 1 - u^{-\alpha}$, and $g(u) = \exp[1 - u^{-\alpha}]$. Therefore, the common gamma frailty model yields the Clayton family of copulas given by Equation B.3:

$$S_{X,Y}(x,y) = \left\{ \frac{1}{S_X(x)^\alpha} + \frac{1}{S_Y(y)^\alpha} - 1 \right\}^{-1/\alpha}, \qquad 0 < \alpha < \infty.$$

This family is particularly useful for constructing bivariate Burr (including Pareto) distributions (see Johnson and Kotz, [13, p. 289]).

EXAMPLE B.5 If the frailty R has a stable distribution with $M_R(z) = \exp\{-(-z)^{1/\alpha}\}$, $\alpha \geq 1$, the corresponding joint survivor function is given by Equation B.4:

$$S_{X,Y}(x,y) = \exp\{-[(-\log S_X(x))^\alpha + (-\log S_Y(y))^\alpha]^{1/\alpha}\}.$$

This family of copulas is particularly useful for constructing bivariate Weibull (including exponential) distributions.

EXAMPLE B.6 If the frailty R has a logarithmic (discrete) distribution on positive integers with $M_R(z) = [\log \alpha]^{-1} \log[1 + e^z(\alpha - 1)]$, then we get the Frank family of copulas given by Equation B.5.

Marshall and Olkin [19] proposed a simulation algorithm for copulas with a frailty construction. This algorithm is applicable to copulas with arbitrary dimensions ($k \geq 2$):

STEP 1 Generate a value r from the random variable R having moment generating function M_R.

STEP 2 Generate independent uniform $(0,1)$ numbers $U_1, \ldots, U_k$.

STEP 3 For $j = 1,\ldots,k$, set $U_j^* = M_R(r^{-1}\log U_j)$, and calculate $X_j = S_j^{-1}(U_j^*)$.

B.4. The Morgenstern Copula

The Morgenstern copula is defined by

$$C(u,v) = uv[1 + \alpha(1-u)(1-v)], \qquad -1 \le \alpha \le 1.$$

This copula cannot be generated by distortion or frailty models.

The following simulation algorithm for the Morgenstern copula can be found in Johnson [17, p. 185]:

STEP 1 Generate independent uniform variables V_1, V_2, and set $U_1 = V_1$.

STEP 2 Calculate $A = \alpha(2U_1 - 1) - 1$ and $B = [1 - \alpha(2U_1 - 1)]^2 + 4\alpha V_2(2U_1 - 1)$.

STEP 3 Set $U_2 = 2V_2/(\sqrt{B} - A)$.

For the Morgenstern copula, Kendall's tau is

$$\tau(\alpha) = \tfrac{2}{9}\alpha, \qquad -1 \le \alpha \le 1,$$

which is limited to the range $(-\frac{2}{9}, \frac{2}{9})$. Thus, the Morgenstern copula can be used only in situations with weak dependence.

An extension of the Morgenstern copula to arbitrary dimensions has been given by Johnson and Kotz [14, 15].

B.5. Summary and Comments

Table B.1 lists the most commonly used bivariate copulas. Except for the reverse monotone copula, they can readily be

TABLE B.1

A SUMMARY OF POPULAR COPULAS

Associated Names	Function Form $C(u,v)$		Kendall's Tau
Independence	uv		0
Common monotone	$\min[u,v]$		1
Reverse monotone	$\max[u+v-1,0]$		-1
Cook–Johnson, Clayton	$[u^{-1/\alpha}+v^{-1/\alpha}-1]^{-\alpha}$	$(\alpha>0)$	$\frac{1}{1+2\alpha}$
	$[u^{-\alpha}+v^{-\alpha}-1]^{-1/\alpha}$	$(\alpha>0)$	$\frac{\alpha}{\alpha+2}$
Frank	$\log_\alpha\left\{1+\frac{(\alpha^u-1)(\alpha^v-1)}{\alpha-1}\right\}$	$(0<\alpha<\infty)$	*
Farlie, Gumbel, Morgenstern	$uv[1+\alpha(1-u)(1-v)]$	$(-1\le\alpha\le 1)$	$\frac{2}{9}\alpha$
Gumbel–Hougaard	$\exp\{-[(-\ln u)^\alpha+(-\ln v)^\alpha]^{1/\alpha}\}$	$(\alpha\ge 1)$	$1-\alpha^{-1}$
normal	$H(\Phi^{-1}(u),\Phi^{-1}(v))$**	$(-1\le\rho\le 1)$	$\frac{2}{\pi}\arcsin(\rho)$

* For the Frank copula,

$$\tau(\alpha)=1+\frac{4}{-\log(\alpha)}\left[\frac{1}{-\log(\alpha)}\int_0^{-\log(\alpha)}\frac{t}{e^t-1}dt-1\right].$$

**H is the joint cumulative distribution function for a bivariate standard normal distribution with a correlation coefficient ρ.
Note that the Cook–Johnson copula with parameter α is the same as the Clayton copula with parameter $1/\alpha$.

extended to multivariate cases ($k>2$). In higher dimensions, the Cook–Johnson copula requires that all taus are the same, while the normal copula allows complete freedom in selecting Kendall's tau.

Some comments on higher dimension extensions of the listed copulas are in order.

1. The independence copula and the common monotone copula both have a unique extension to higher dimensions, while the reverse monotone has multiple possible extensions.

2. Recall that the Cook–Johnson copula, the Clayton copula, the Frank copula, and the Gumbel–Hougaard copula can be generated by the distortion method. They can be readily generated to higher dimensions by the relation

$$g[S_{X_1,X_2,\ldots,X_k}(t_1,t_2,\ldots,t_k)] = g[S_{X_1}(t_1)] \cdot g[S_{X_2}(t_2)] \cdots g[S_{X_k}(t_k)].$$

   However, the correlation structure is restricted in a sense that Kendall's tau (or rank correlation coefficient) is the same for any pair of variables.

3. The Morgenstern copula can be generated to higher dimensions, but the parameter values are further restricted.

4. The normal copula stands out among others for its extremely flexible correlation structure at higher dimensions. It allows complete freedom in selecting Kendall's taus or rank order coefficients, as we have seen in Section 8.

Frees and Valdez [6] have written a good survey paper on the use of copulas, including the associated simulation techniques. In general, the use of copulas permits simple implementation by Monte Carlo simulation, thus aggregating correlated risks.

B.6. The Use of Mixed Copulas

Assume that joint cumulative distribution functions $F_{X_1,\ldots,X_k}$ and $G_{X_1,\ldots,X_k}$ have the same marginals $F_{X_1},\ldots,F_{X_k}$. Then the mixed joint cumulative distribution function

$$(1-q)F_{X_1,\ldots,X_k}(t_1,\ldots,t_k) + qG_{X_1,\ldots,X_k}(t_1,\ldots,t_k), \qquad 0 < q < 1,$$

also has a marginal cumulative distribution function $F_{X_1},\ldots,F_{X_k}$. For this mixed joint distribution, we have

$$\tau[X_i,X_j] = (1-q)\tau^F[X_i,X_j] + q\tau^G[X_i,X_j],$$

where τ^F and τ^G represent Kendall's taus implied by the joint cumulative distribution functions F and G, respectively. In particular, if we let $F_{X_1,\ldots,X_k}(t_1,\ldots,t_k) = \prod_{j=1}^{k} t_j$ represent the independent copula and $G_{X_1,\ldots,X_k}(t_1,\ldots,t_k) = \min[t_1,\ldots,t_k]$ represent the comonotonic copula, then $\tau[X_i,X_j] = q$.

The mixture of joint cumulative distribution functions can be used to adjust, up or down, Kendall's tau. For example, if we feel that a common mixture joint cumulative distribution function F would give too strong a correlation, then we can mix it with an independent joint cumulative distribution function G. If we feel that a common mixture joint cumulative distribution function F would give too weak a correlation, then we can mix it with a comonotonic joint cumulative distribution function G'.

APPENDIX C

PANJER'S RECURSIVE METHOD

As an alternative to the FFT method, we introduce Panjer's recursive method for evaluation of aggregate loss distributions.

Suppose that the severity distribution $f_X(x)$ is defined on $0,1,2,\ldots$, representing multiples of some convenient monetary unit.

Suppose that the frequency distribution is a member of the (a,b) class satisfying

$$\Pr\{N=k\} = \left(a + \frac{b}{k}\right)\Pr\{N=k-1\}, \qquad k = 1,2,3,\ldots . \tag{C.1}$$

Note that the Poisson and negative binomial distributions are included in this family.

For the Poisson distribution, we have $a = 0$ and $b = \lambda$.

For the negative binomial (α,β) distribution, we have

$$a = \frac{\beta}{1+\beta} \qquad \text{and} \qquad b = \frac{(\alpha-1)\beta}{1+\beta}.$$

Panjer [21] has shown that the aggregate loss distribution $f_S(x)$ can be evaluated recursively

$$f_S(x) = \left[\sum_{y=1}^{x}\left(a + \frac{by}{x}\right) f_X(y) f_S(x-y)\right](1 - a f_X(0))^{-1}. \tag{C.2}$$

The starting value of the recursive algorithm is $f_S(0) = P_N(f_X(0))$.

In the case of the Poisson distribution, it reduces to

$$f_S(x) = \frac{\lambda}{x}\sum_{y=1}^{x} y f_X(y) f_S(x-y), \qquad x = 1,2,\ldots, \tag{C.3}$$

with starting value

$$f_S(0) = e^{-\lambda(1-f_X(0))}. \tag{C.4}$$

The recursive method is fairly easy to program.

For example, suppose that the claim frequency N has a Poisson distribution with mean $\lambda = 3$ and the claim severity X has a probability distribution

$$\Pr\{X = 1\} = 0.5, \qquad \Pr\{X = 2\} = 0.3, \qquad \Pr\{X = 3\} = 0.2, \tag{C.5}$$

then the probability distribution of the aggregate loss S can be calculated recursively

$$f_S(0) = e^{-\lambda} = 0.04979,$$

$$f_S(1) = \frac{\lambda}{1}[f_X(1)f_S(0)] = 0.07468,$$

$$f_S(2) = \frac{\lambda}{2}[f_X(1)f_S(1) + f_X(2)f_S(0)] = 0.07842,$$

$$f_S(3) = \frac{\lambda}{3}[f_X(1)f_S(2) + f_X(2)f_S(1) + f_X(3)f_S(0)] = 0.07157.$$

APPENDIX D

SOME FREQUENTLY ASKED QUESTIONS

Q1. In this paper a lot of discussion has been given to correlated frequency models. What about models of correlation between claim severities?

A1. We have a general correlation model—normal copula—which can be used to model correlated claim severities. In fact, @Risk (which is an Excel add-in application) can be readily used to carry out such simulations. But one should keep in mind that the correlation parameters used in @Risk are rank correlation coefficients.

Correlation in claim severities often comes from the uncertainty in the future claim inflation and loss development. A simple method of quantifying this correlation is to use a common multiplier:

$$X_1 = B \cdot Y_1, \ldots, X_j = B \cdot Y_j, \ldots X_N = B \cdot Y_N,$$

where

- the Y_js are independent,
- the number of claims N may be fixed or random but independent from the severity Y_js and the multiplier B, and
- the common multiplier B may be assigned a probability distribution (discrete or continuous).

The sum of the k losses is

$$Z = X_1 + X_2 + \cdots + X_N = B \cdot (Y_1 + Y_2 + \cdots + Y_N).$$

Thus, one can first evaluate the independent sum of the Y_js, and then combine it with the multiplier B. In this model, we have $\mathrm{Cov}[X_i, X_j] = \mathrm{Var}(B) \cdot \mathrm{E}[Y_i] \cdot \mathrm{E}[Y_j]$.

Q2. In some situations, the claim frequency and severity are correlated. How would one construct such a model?

A2. In some catastrophe modeling it might be plausible to consider the dependency between claim frequency and claim severity. For instance, the Richter scale value of an earthquake may affect the claim frequency and severity simultaneously, and for hurricane losses, the wind speed would affect both the claim frequency and severity in the same direction. For a modeling of such catastrophe losses, a good understanding of the underlying loss generating mechanism is essential, which requires sound knowledge in meteorology, construction engineering, population density, insurance coverage, etc. Some have observed that demand surge after a major catastrophe may also generate correlation between claim frequency and severity. Nevertheless, mathematics can serve as a tool to quantify our understanding of the underlying loss generating mechanism.

For reinsurance excess-of-loss modeling, the frequency and severity of a given layer may be positively correlated in a high inflation environment. This is due to the leverage effect of inflation. This correlation can be quantified by using a random trending factor for ground-up losses and then quantifying the frequency/severity for losses for a reinsurance layer.