1 Modelling II: Conditional and non-linear

1.1 Introduction

Insurance requires modelling tools different from those of the preceding chapter. Pension schemes and life insurance make use of **life cycle** descriptions. Individuals start as 'active' (paying contributions), at one point they 'retire' (drawing benefits) or become 'disabled' (benefits again) and they may die. To keep track on what happens stochastic models are needed, but those can not be constructed by means of linear relationships like in the preceding chapter. There are no numerical variables to connect! Distributions are used instead.

The central concept is **conditional** probabilities and distributions, expressing mathematically that what has occurred is going to influence (but not determine) what comes next. That idea is the principal topic of the chapter. As elsewhere, mathematical aspects (here going rather deep) are downplayed. Our target is the conditional viewpoint as a modelling tool. Sequences of states in life cycles involve time series (but of kind different from those in Chapter 5) and are treated in Section 6.6. Actually time may not be involved at all. Risk hetereogenity in property insurance is a typical (and important) example. Consider a car owner. What he encounters daily in the traffic is thoroughly influenced by randomness, but so is (from a company point of view) his ability as a driver. These are uncertainties of entirely different origin and define a **hierarchy** (driver comes first). Conditional modelling is the natural way to connect random effects of this kind which operate on different levels. The very same viewpoint is used when estimation and Monte Carlo errors are examined in the next chapter, and there are countless other examples.

Conditional arguments will hang over much of this chapter, and we embark on it in the next section. **Copulas** is an additional tool. The idea behind is very different from conditioning and as a popular approach of fairly recent origin. Yet copulas has without doubt to come to stay. Section 6.7 is an introduction.

1.2 Conditional modelling

Introduction

Condional probabilities are defined in elementary textbooks in statistics. When an event A has occurred, the probability of another one B changes from $\Pr(B)$ to

$$\Pr(B|A) = \frac{\Pr(A \cap B)}{\Pr(A)} \qquad \text{and also} \qquad \Pr(B|A) = \frac{\Pr(A|B)\Pr(\mathrm{B})}{\Pr(A)}, \tag{1.1}$$

where the right hand side is known as **Bayes' formula**. This is of obvious relevance in gambling where new information leads to new odds. In this book conditional probabilities are above all modelling tools, used to express random relationships between random variables. Note the mathematical notation. The condition is always placed to the right of a vertical bar. Similar notation is $Y|x$ (the random variable Y given x) with conditional density functions and expectaton $f(y|x)$ and $E(Y|x)$.

Conditional modelling is **sequential** modelling, first X and then Y given X. The purpose of this section is to demonstrate the power in this line of thinking. It is the natural way to describe countless stochastic phenomena, and simulation is easy. Simply

generate X^* and then Y^* given X^*,

the second drawing being dependent on the outcome of the first. The following examples all play major roles in later chapters.

The conditional Gaussian

Bivariate normal models were in Chapter 2 defined through

$$X_1 = \xi_1 + \sigma_1\eta_1 \qquad \text{and} \qquad X_2 = \xi_2 + \sigma_2(\rho\eta_1 + \sqrt{1-\rho^2}\,\eta_2),$$

where η_1 and η_2 are independent and $N(0,1)$; see (**??**). Suppose $X_1 = x_1$ is fixed. Then $\eta_1 = (x_1 - \xi_1)/\sigma_1$, which when inserted for η_1 in the representation for X_2 leads to

$$X_2 = \xi_2 + \sigma_2(\rho\frac{x_1-\xi_1}{\sigma_1} + \sqrt{1-\rho^2}\eta_2),$$

or after some reorganizing

$$X_2 = \underbrace{(\xi_2 + \rho\sigma_2\tfrac{x_1-\xi_1}{\sigma_1})}_{\textit{expectation}} + \underbrace{(\sigma_2\sqrt{1-\rho^2})}_{\textit{standard deviation}}\eta_2. \tag{1.2}$$

Here η_2 is the only random term and, by definition, X_2 is normal with mean and standard deviation

$$E(X_2|x_1) = \xi_2 + \rho\sigma_2\frac{x_1-\xi_1}{\sigma_1} \qquad \text{and} \qquad \text{sd}(X_2|x_1) = \sigma_2\sqrt{1-\rho^2}. \tag{1.3}$$

We are dealing with a **conditional distribution**. As x_1 is varied, then so does the expectation and (for other models) also standard deviation.

Survival modelling

Let Y be the length of life of an individual. A central quantity in life insurance is

$${}_tp_{y_0} = \Pr(Y > y_0 + t | Y > y_0), \tag{1.4}$$

known as the **survival probability**. This defines how likely it is that a person of age y_0 reaches age $y_0 + t$. If $F(y)$ is the distribution function of Y, then from (1.1) left

$${}_tp_{y_0} = \frac{\Pr(Y > y_0 + t)}{\Pr(Y > y_0)} = \frac{1 - F(y_0 + t)}{1 - F(y_0)} \qquad \text{for} \qquad y_0,\, t > 0. \tag{1.5}$$

Survival probabilities are often used on multiples of a given increment h, for example

$$\underbrace{y_l = lh}_{\textit{age}} \quad l = 0, 1\ldots \qquad \text{and} \qquad \underbrace{t_k = kh}_{\textit{time}} \quad k = 0, 1\ldots,$$

and we shall write ${}_kp_{l_0} = {}_tp_{y_0}$ when $y_0 = l_0h$ and $t = kh$ and also $p_l = {}_1p_l$. The probability of surviving the coming k time steps must be equal to

$${}_kp_{l_0} = \underbrace{p_{l_0}}_{\textit{first interval}} \times \underbrace{p_{l_0+1}}_{\textit{second interval}} \times \cdots \times \underbrace{p_{l_0+k-1}}_{\textit{k'th interval}}, \tag{1.6}$$

and survival modelling is built up from the one-step probabilities p_l; more on that in Chapter 12.

Over threshold modelling

Conditional probabilities of the same type is needed in property insurance, particularly in connection with large claims and re-insurance. For a given threshold b we seek the distribution of

$$Z_b = Z - b \qquad \text{given that} \qquad Z > b. \tag{1.7}$$

We can write it down by replacing t and y_0 on the right in (1.5) by z and b. Thus

$$\Pr(Z_b > z | Z > b) = \frac{1 - F(b+z)}{1 - F(b)},$$

where $F(z)$ is the distribution function of Z. When differentiated with respect to z, this leads to

$$f_b(z) = \frac{f(z+b)}{1 - F(b)}, \qquad z > 0. \tag{1.8}$$

as the density function for the amount exceeding a given threshold. Tail distributions of this type possess a remarkable property, see Pickands (1975). For most distributions used in practice, precisely if $f(z)$ is *not* identically zero above some upper limit, then $f_b(z)$ becomes either a *Pareto* density or an *exponential* one as $b \to \infty$ no matter which model we started with; see Chapter 9.

Risk hetereogenity

Claim numbers N in property insurance was in Chapter 3 described by

$$\Pr(N = n|\mu) = \frac{\lambda^n}{n!}\exp(-\lambda) \qquad \text{where} \qquad \underset{\textit{Policy}}{\lambda = \mu T} \qquad \text{or} \qquad \underset{\textit{Portfolio}}{\lambda = J\mu T},$$

for $n = 0, 1, \ldots$; see (**??**) and (**??**). The nomenclature implies that μ now has become random, and there are many situations where this is a natural viewpoint. Consider automobile insurance. Variation in risk among drivers is accounted for by one random μ for each individual, but μ might also reflect general driving conditions (such as the weather) that affect everybody jointly.

Modelling is the same whether μ affects policies *individually* or the entire portfolio *collectively.* The claim frequency observed (N for an individual or $\mathcal{N}$ for a portfolio) is the outcome of two experiments in a hierarchy. First μ is drawn randomly and *then* N or $\mathcal{N}$ given μ; i.e.

$$\underset{\textit{policy level}}{\mu = \xi Z, \quad N|\mu \sim \text{Poisson}(\mu T)} \qquad \text{and} \qquad \underset{\textit{portfolio level},}{\mu = \xi Z, \quad \mathcal{N}|\mu \sim \text{Poisson}(J\mu T).} \tag{1.9}$$

where $E(Z) = 1$ to make ξ *mean intensity.* The standard model for Z is Gamma(α), one of the distributions introduced in Section 2.6. Now

$$E(\mu) = \xi \qquad \text{and} \qquad \text{sd}(\mu) = \xi/\sqrt{\alpha}, \tag{1.10}$$

and the variability in μ, controlled by α, is removed when $\alpha \to \infty$ so that eventually $\mu = \xi$ becomes fixed; more on this model in Section 8.3.

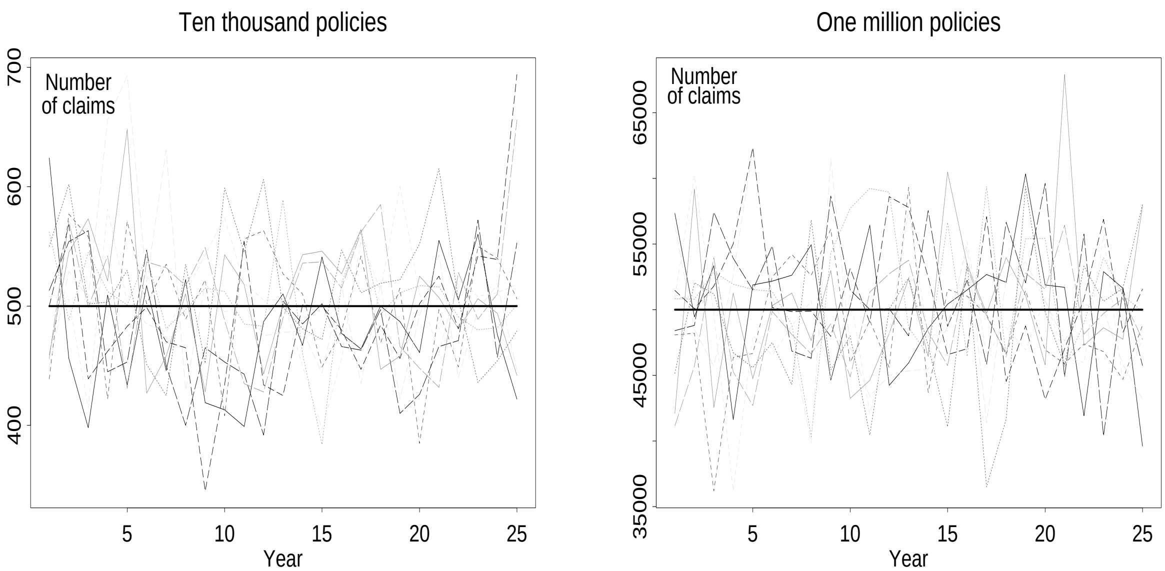


Figure 6.1 *Simulated portfolio claim frequency scenarios under annual change of risk*

Common risk factors

Claim numbers $N_1, \ldots, N_J$ depending on the same random intensity μ is a special case of a more general viewpoint. A random variable ω is called a **common factor** for $X_1, \ldots, X_J$ if

$$X_1, \ldots, X_J \qquad \text{are } \textbf{conditionally} \text{ independent} \qquad \text{given } \omega. \tag{1.11}$$

The theory of crdibility in Section 10.5 is a classical example in actuarial science, and the market component played the role of ω in the CAPM model of Section 5.3. If ω isn't directly observable, we are dealing with **hidden** or **latent** factors.

Common factors (whether hidden or not) invariably increase risk and they are impossible to diversify. Figure 6.1 is a simulated example where claim frequencies were generated over 25 years for one 'small' and one 'large' car insurance portfolio. The intensity μ, changed every year and was the same for all policies. Suppose μ follows a Gamma model. Claim frequencies are then generated through

$$Z^* \sim \text{Gamma}(\alpha), \qquad \mu^* \leftarrow \xi Z^* \qquad \text{and then} \qquad \mathcal{N}^* \sim \text{Poisson}(J\mu^* T).$$

The experiments in Figure 6.1 were run as 25 independent drawings for each of $m = 20$ scenarios plotted jointly. Underlying parameters were

$$\xi = 5\%, \qquad \alpha = 100, \qquad T = 1,$$

which means that claim frequency per car is 5% in an average year and the standard deviation 10% of that; see (1.10). Fluctuations in Figure 6.1 match this fairly well[1], but the main point is the

[1]The oscillations in both plots go out to about $\pm 20\%$ of the position of the straight line, and the 10% relative standard deviation emerges when you divide by two.

uncertainty which is no smaller (in relative terms) for the large portfolio. That runs contrary to what has seen before (Section 3.2) and reflects that the impact of common factors isn't removed through portfolio size. The mathematics is given in Section 6.3.

Monte Carlo distributions

Simulation experiments are often run from parameters that have been estimated from historical data. The distribution of the simulations are then influenced by estimation error *in addition* to ordinary Monte Carlo randomness. To be specific, suppose the claim frequency $\mathcal{N}$ against a portfolio follows the ordinary Poisson model with fixed intensity μ and let $\hat{\mu}$ be its estimate (method in Chapter 8). The scheme is then

$$\text{historical data} \underset{\textit{estimation}}{\longrightarrow} \hat{\mu} \underset{\textit{Monte Carlo}}{\longrightarrow} \mathcal{N}^*,$$

and the question is how both sources of error are aggregated. A first step is to notice that the model for $\mathcal{N}^*$ is really a *conditional* one; i.e

$$\Pr(\mathcal{N}^* = n|\hat{\mu}) = \frac{(JT\hat{\mu})^n}{n!}\exp(-JT\hat{\mu}), \qquad n = 0, 1, \ldots,$$

and we must combine with statistical errors in the estimation process. The argument is given in Chapter 7.

1.3 Risk from subordinate level

Introduction

Risk variables X influenced by a second random factor ω on a **subordinate** level were introduced above. Think of ω as personal qualities of a policy holder, background conditions affecting an entire insurance portfolio or market risk in finance. The importance of this kind of uncertainty was examined in Section 5.3 through a specific model (CAPM), but it is also possible to proceed more generally through the conditional mean and standard deviation only. To this end let

$$\xi(\omega) = E(X|\omega) \qquad \text{and} \qquad \sigma(\omega) = \text{sd}(X|\omega), \tag{1.12}$$

and the aim of this section is to examine the risk of an insurance portfolio when mean and standard deviation of individual claims vary with ω.

The double rules

Our tool is two operational rules that is best introduced generally. Suppose the distribution of Y depends on a random vector $\mathbf{X}$. The condional mean $\xi(\mathbf{x}) = E(Y|\mathbf{x})$ plays a leading role in risk modelling; see Section 6.4. Here the issue is how $\xi(\mathbf{x})$ and $\sigma(\mathbf{x}) = \text{sd}(Y|\mathbf{x})$ influence Y. Much insight is provided by the identities

$$\underset{\textit{double expectation}}{E(Y) = E\{\xi(\mathbf{X})\}} \qquad \text{for} \qquad \xi(\mathbf{x}) = E(Y|\mathbf{x}) \tag{1.13}$$

and

$$\underset{\textit{double variance}}{\text{var}(Y) = \text{var}\{\xi(\mathbf{X})\} + E\{\sigma^2(\mathbf{X})\}} \qquad \text{for} \qquad \sigma(\mathbf{x}) = \text{sd}(Y|\mathbf{x}), \tag{1.14}$$

which are proved in Appendix A. Both $\xi(\mathbf{X})$ and $\sigma^2(\mathbf{X})$ are random variables, and the expectation of the former is the expectation of Y. The double variance formula decomposes $\mathrm{var}(Y)$ into two positive contributions with consequences reaching far.

Neither formula requires conditional modelling beyond mean and standard deviation, and a number of useful results can be derived from them. They will in Chapter 7 play a main role when errors of different origin are analysed jointly.

Impact of subordinate risk
Let X_j be a random variable depending on ω_j for $j = 1, \dots, J$. The independence of $X_1, \dots, X_J$ is still assumed, but this now means that they are **conditionally independent** given $\omega_1, \dots, \omega_J$. Consider the two different sampling regimes

$$\underbrace{\omega_1 = \ldots = \omega_J = \omega}_{\textit{common factor}} \qquad \text{and} \qquad \underbrace{\omega_1, \ldots, \omega_J \quad \text{independent.}}_{\textit{individual parameters}}$$

On the left ω is a common background factor affecting the entire portfolio whereas on the right ω_j is attached each X_j individually. Their effect on the portfolio risk $\mathcal{X} = X_1 + \ldots + X_J$ is widely different, as we shall now see. It will be assumed that all X_j and all ω_j follow the same distribution (*not* essential).

Consider first the case where ω is a common background factor for the entire portfolio. We are assuming that $\xi(\omega) = E(X_j|\omega)$ and $\sigma(\omega) = \mathrm{sd}(X_j|\omega)$, the same for all j. Hence, by adding all contributions

$$E(\mathcal{X}|\omega) = J\xi(\omega) \qquad \text{and} \qquad \mathrm{var}(\mathcal{X}|\omega) = J\sigma^2(\omega),$$

where the variance formula requires conditional independence. Invoke the double rules with $Y = \mathcal{X}$ and $\mathbf{X} = \omega$. Then, by (1.14)

$$\mathrm{var}(\mathcal{X}) = \mathrm{var}\{J\xi(\omega)\} + E\{J\sigma^2(\omega)\} = J^2\mathrm{var}\{\xi(\omega)\} + JE\{\sigma^2(\omega)\}$$

which with (1.13) leads to

$$E(\mathcal{X}) = JE\{\xi(\omega)\} \qquad \text{and} \qquad \mathrm{sd}(\mathcal{X}) = \underbrace{J\sqrt{\mathrm{var}\{\xi(\omega)\} + E\{\sigma^2(\omega)\}/J}}_{\textit{common } \omega}, \tag{1.15}$$

and the standard deviation is of the same order of magnitude J as the expectation itself. *Such risk can not be diversified away* by increasing the portfolio size. Indeed,

$$\frac{\mathrm{sd}(\mathcal{X})}{E(\mathcal{X})} \to \frac{\mathrm{sd}\{\xi(\omega)\}}{E\{\xi(\omega)\}} \qquad \text{as} \qquad J \to \infty$$

which does not vanish if $\mathrm{sd}\{\xi(\omega)\} > 0$.

Things change drastically when each X_j is attached a separate and independently drawn ω_j. The mean and variance of each X_j are now calculated by inserting $J = 1$ in (1.15). When all of those are added over all policies j, we obtain mean and variance on portfolio level; i.e

$$E(\mathcal{X}) = JE\{\xi(\omega)\} \qquad \text{and} \qquad \mathrm{sd}(\mathcal{X}) = \underbrace{\sqrt{J[E\{\sigma^2(\omega)\} + \mathrm{var}\{\xi(\omega)\}]}}_{\omega \textit{ individual}}. \tag{1.16}$$

The mean is the same as before, but the standard deviation has now the familiar form proportional to $\sqrt{J}$.

Example: Random claim intensity.
The preceding argument enables us to understand how random intensities $\mu_1 \ldots, \mu_J$ influence the claim frequency $\mathcal{N} = N_1 + \ldots + N_J$ of the portfolio under the two sampling regimes above:

$$\underbrace{\mu_1 = \ldots = \mu_J = \mu}_{\textit{common factor}} \qquad \text{and} \qquad \underbrace{\mu_1, \ldots, \mu_J \quad \text{independent.}}_{\textit{individual parameters}}$$

On the left a common (random) factor μ is allocated all policy holders jointly whereas on the right there is one independent intensity for each individual. Claim frequencies $N_1, \ldots, N_J$ are in either case assumed conditionally independent and Poisson given $\mu_1, \ldots, \mu_J$. In particular

$$E(N_j|\mu_j) = \mu_j T \qquad \text{and} \qquad \text{var}(N|\mu_j) = \mu_j T.$$

which are the functions $\xi(\mu_j)$ and $\sigma^2(\mu_j)$ in (1.15) and (1.16). Let $\xi_\mu = E(\mu_j)$ and $\sigma_\mu = \text{sd}(\mu_j)$ be the mean and standard deviation of each μ_j. Then (1.15) yields

$$E(\mathcal{N}) = JT\xi_\mu \qquad \text{and} \qquad \text{sd}(\mathcal{N}) = \underbrace{JT\sqrt{\sigma_\mu^2 + \xi_\mu/(JT)}}_{\textit{common } \mu}, \tag{1.17}$$

and the form of the standard deviation (almost proportional to J) explains the simulated patterns in Figure 6.1 where the relative random uncertainty seemed unaffected by J.

This changes when $\mu_1, \ldots, \mu_J$ are drawn independently of each other. Now

$$E(\mathcal{N}) = JT\xi_\mu \qquad \text{and} \qquad \text{sd}(\mathcal{N}) = \underbrace{T\sqrt{J(\sigma_\mu^2 + \xi_\mu/T)}}_{\mu \textit{ individual}}, \tag{1.18}$$

and the standard deviation has the familiar form proportional to $\sqrt{J}$. The practical significance for portfolio risk will be examined below.

Insurance risk: A simple formula
Another useful consequences of the double rules are simple formulae for mean and standard deviation of total portfolio loss. Consider the model from Section 3.2; i.e

$$\mathcal{X} = Z_1 + \ldots + Z_\mathcal{N}$$

where $\mathcal{N}, Z_1, Z_2 \ldots$ are stochastically independent. Let $E(Z_i) = \xi_z$ and $\text{sd}(Z_i) = \sigma_z$. Elementary rules for expectation and variance of sums yields

$$E(\mathcal{X}|\mathcal{N}) = \mathcal{N}\xi_z \qquad \text{and} \qquad \text{var}(\mathcal{X}|\mathcal{N}) = \mathcal{N}\sigma_z^2.$$

To incorporate claim frequency $\mathcal{N}$ as an additional source of randomness take $Y = \mathcal{X}$ and $\mathbf{X} = \mathcal{N}$ in (1.13) and (1.14). Then

$$\text{var}(\mathcal{X}) = \text{var}(\mathcal{N}\xi_z) + E(\mathcal{N}\sigma_z^2) = \text{var}(\mathcal{N})\xi_z^2 + E(\mathcal{N})\sigma_z^2$$

so that

$$E(\mathcal{X}) = E(\mathcal{N})\xi_z \qquad \text{and} \qquad \text{var}(\mathcal{X}) = E(\mathcal{N})\sigma_z^2 + \text{var}(\mathcal{N})\xi_z^2. \tag{1.19}$$

In particular, suppose $\mathcal{N}$ follows a pure Poisson distribution. Then $E(\mathcal{N}) = \text{var}(\mathcal{N}) = J\mu T$ and

$$E(\mathcal{X}) = J\mu T\xi_z \qquad \text{and} \qquad \text{var}(\mathcal{X}) = J\mu T(\sigma_z^2 + \xi_z^2), \tag{1.20}$$

which will be used repeatedly.

Random claim intensity: Important at portfolio level?
How the portfolio liabilities $\mathcal{X}$ are affected by random claim intensities may be examined by inserting the expressions for $E(\mathcal{N})$ and $\text{sd}(\mathcal{N})$ into (1.19). For the mean this leads to

$$E(\mathcal{X}) = J\xi_\mu\xi_z, \tag{1.21}$$

which applies both when μ is a common, random value for the entire portfolio and when drawn individually for each policy holder. This changes when we move to the variance. Inserting (1.17) and (1.18) into (1.19) right yields by some algebra (detailed in Section 6.8)

$$\text{sd}(\mathcal{X}) = \underbrace{\sqrt{J\xi_\mu(\sigma_z^2 + \xi_z^2)}}_{\textit{pure Poisson}} \times \underbrace{\sqrt{1+\delta\gamma}}_{\textit{due to random } \mu} \tag{1.22}$$

where

$$\delta = T\frac{\sigma_z^2}{\sigma_z^2 + \xi_z^2} \cdot \frac{\sigma_\mu^2}{\xi_\mu} \qquad \text{and} \qquad \gamma = \begin{matrix} 1 & \text{for } \textbf{indvidual } \mu \\ J & \text{for } \textbf{common } \mu \end{matrix} \tag{1.23}$$

This lengthy expression tells a lot.

On the right in (1.22) there is a main, pure Poisson factor and a correction caused by μ being random. How important is the latter? In practice δ is quite small (hardly more than a few per cent, see Exercise 6.3.2), and when $\mu_1, \dots, \mu_J$ are drawn independently of each other (so that $\gamma = 1$), the correction factor becomes $\sqrt{1+\delta} \doteq 1 + \delta/2$, *not* a large increase. The other case is different. Now $\gamma = J$, and the correction $\sqrt{1+J\delta}$ may be huge.

1.4 The role of the conditional mean

Introduction
The conditional mean is much more than a brick in the double rules of the preceding section. Consider

$$\hat{Y} = \xi(\mathbf{X}) = E(Y|\mathbf{X}), \tag{1.24}$$

where $\mathbf{X}$ is a quantity observed. In theory $\hat{Y}$ is the best way of predicting the value of an unknown Y if you know $\mathbf{X}$. This is a celebrated result in engineering and statistics, yet not *that* prominent in actuarial science. When Y is a future value, we are often more concerned with summaries such as mean and percentiles than with predicting its actual outcome.

But there is another (and important) side to this. If $\mathbf{X}$ is information available, $E(Y|\mathbf{X})$ is what is expected given that knowledge and a natural break-even price for carrying the risk Y. Shouldn't what we charge reflect what we know? In property insurance $\mathbf{X}$ might be our experience with a policy holder or other information with bearing on risk; see Part II. Here the main example is the pricing of money market products such as bonds which will eventually lead us to the theoretical interest rate curve.

A quick word on the meaning of $\mathbf{X}$ in the present context: Think of it as all present and past observations with bearing on Y. Theoretical literature often refers to $\mathbf{X}$ as a **sigma-field** (typically denoted $\mathcal{F}$), but it is perfectly possible to understand the ideas involved without such formalism from measure theory.

Optimal prediction and interest rates
Central mathematical properties of the conditional mean are

$$\underbrace{E(\hat{Y}-Y)=0}_{\textit{expected error}} \qquad \text{and} \qquad \underbrace{E(\hat{Y}-Y)^2 \leq E(\tilde{Y}-Y)^2}_{\textit{expected squared error}} \quad \text{for all } \tilde{Y}=\tilde{Y}(\mathbf{X}). \tag{1.25}$$

Here the left hand side, which is merely a rephrasal of the rule of double expectation (1.13), signifies that the expected prediction error $\hat{Y}-Y$ is zero. This means that $\hat{Y}$ is an **unbiased** prediction; more on unbiasedness in Chapter 7. The inequality on the right in (1.25) shows that the conditional mean is on average the most accurate way of utilizing the information $\mathbf{X}$. The proof is given in Section 6.8.

This result will now be used to examine interest rate forecasting. Suppose the rates $\{r_k\}$ follow the Vasicěk model of Section 5.6. Then

$$r_k = \xi + \sigma(\varepsilon_k + a\varepsilon_{k-1} + \ldots + a^{k-1}\varepsilon_1) + a^k(r_0 - \xi),$$

where r_0 is known and $\varepsilon_1, \varepsilon_2, \ldots$ are independent with zero mean and unit variance; see (**??**). It follows that

$$E(r_k|r_0) = \xi + a^k(r_0 - \xi) \qquad \text{and} \qquad \text{sd}(r_k|r_0) = \sigma\sqrt{\frac{1-a^{2k}}{1-a^2}};$$

see Section 5.6 for the standard deviation. If the Vasicěk model is true, $\hat{r}_k = E(r_k|r_0)$ is the best possible prediction of r_k.

What about the accuracy? A quick look is provided by the formula for the standard deviation. Possible annual parameters *could be* $\sigma = 0.016$ and $a = 0.7$. If so, the standard deviation becomes 1.4% after one year and 2.2% after five. This signifies huge prediction error, up to $3-4\%$ and more. Forecasting interest levels though simple statistical techniques is futile.

Term structure modelling
The conditional mean is in the money market much more important for pricing than for prediction. As above let r_0 be the rate of interest today (known) and $r_1, \ldots, r_k$ those of the future (unknown), running over the time sequence $t_k = kh$ for $k = 0, 1, \ldots$. Consider

$$P(r_0, t_k) = E_Q(D_k|r_0) \qquad \text{where} \qquad D_k = \frac{1}{1+r_1}\cdots\frac{1}{1+r_k}. \tag{1.26}$$

Here D_k is a discount and had future rates of interest been known, that is what you would have been charged today for the right to receive one money unit at t_k; i.e. for a unit-faced, zero-coupon bond expiring at that time. The rates $r_1, \ldots, r_k$ are (of course) unknown, but there are views on what they are going to be, leading to the expectation of D_k on the left in (1.26). This is a theoretical price of a zero-coupon bond. When the situation is analysed in Chapter 14, it will emerge that a **risk-neutral** Q-model should be employed, hence the subscript Q. This Q-model is *not* the same as the one describing real interest rate fluctuations. It was the same with equity option in Section 3.5.

There are now *two* pricing systems for zero-coupon bonds. The preceding theoretical **term structure** $P(r_0, t_k)$ is a based on a mathematical description of the market view, but there are also the observed prices $P(0\!:\!k)$ introduced in Section 1.4. Why bother with the theoretical ones at all? Answer: We need them to describe *future* bond prices and their uncertainty. For example, suppose r_k^* is a Monte Carlo rate of interest at t_k. Then $P(r_k^*, t)$ is a simulated price at time t_k of a bond expiring at $t_k + t$. Such simulations will in Chapter 15 play a crucial role with modern **fair value** accounting and with the coordination of assets and liabilities in the life insurance industry.

Surely the observed bond prices $P(0\!:\!k)$ and the theoretical ones $P(r_0, t_k)$ at time $t_0 = 0$ should be equal? They must if the Q-model correctly reflects the market view, and it is common to calibrate parameters by matching the two sets of prices.

Example: The Vasicěk term structure
Countless theoretical bond pricing schemes have appeared in the literature; see Section 6.9. They make use of a mathematical limiting process where $h \to 0$ and carefully constructed Q-models that allow explicit formulae. One of the simplest and most widely used is the Vasicěk model

$$r_k - r_{k-1} = a_q h(\xi_q - r_{k-1}) + \sqrt{h}\sigma_q \varepsilon_k,$$

which was introduced through (**??**). The parameters are now subscripted with q to emphasize risk-neutrality. Calculations of (1.26) under this model are carried out in Exercises 5.7.12-16 which lead to

$$P(r_0, t) = e^{A(t) - B(t) r_0} \tag{1.27}$$

where

$$B(t) = \frac{1 - e^{-a_q t}}{a_q} \qquad \text{and} \qquad A(t) = (B(t) - t)\left(\xi_q - \frac{\sigma_q^2}{2a_q^2}\right) - \frac{\sigma_q^2 B(t)^2}{4a_q}. \tag{1.28}$$

We may interprete $P(r_0, t)$ as the price in a Vasicěk world of a zero-coupon bond maturing at time t when r_0 is the present rate of interest. The model goes back to Vasicěk (1977).

Monte Carlo term structures
With modern computational power simple bond price formulae may not be so important as before, and it is perfectly feasible to compute $P(r_0, t_k)$ by Monte Carlo and store it as a table over a suitable set of pairs (r_0, t_k). Simulations such as $P(r_k^*, t)$ may then be read off approximately from the table. The minor numerical inaccuracy is of little practical importance, and we may now employ

any Q-model we like. Here is an implementation for the Black-Karisinsky model (which does not allow simple bond price schemes):

Algorithm 6.1 The Black-Karisinsky term structure

Input: m, ξ_q, a_q, σ_q, r_0, h and
$\quad\sigma_x = \sigma_q/\sqrt{1-a_q^2}$, $x_0 = \log(r_0/\xi_q) + \sigma_x^2/2$
$P^*(k) \leftarrow 0$ for $k = 1, \ldots, K$ *%$P^*(k)$ the theoretical bond price*
Repeat m times
$\quad X^* \leftarrow x_0$, $D^* \leftarrow 1/m$ *%D^* will serve as discount*
$\quad$For $k = 1, \ldots, K$ do
$\quad\quad$Draw $\varepsilon^* \sim N(0,1)$ and $X^* \leftarrow a_q X^* + \sigma_q \varepsilon^*$
$\quad\quad r^* \leftarrow \xi_q e^{-\sigma_x^2/2+X^*}$ and $D^* \leftarrow D^*/(1+r^*)$
$\quad\quad P^*(k) \leftarrow P^*(k) + D^*$ *%The k-step discount summarized*

Return $P^*(k)$ for $k = 1, \ldots, K$

The algorithm simulates future rates of interest and updates the stochastic discounts as it goes through the *inner* loop over k. Output from the *outer* loop are Monte Carlo approximations $P^*(k)$ to $P(r_0, t_k)$ for $k = 1, \ldots, K$. Re-runs for many different r_0 are necessary.

If you want the computations to run on a finely meshed time scale, you must adapt the parameters as explained in Section 5.7. The examples in Figure 6.2 have been run om a crude annual one with parameters

$$\xi_q = 4\%, \quad a_q = 0.7, \quad \sigma_q = 0.25 \qquad \text{or} \qquad \xi_q = 4\%, \quad a_q = 0.5, \quad \sigma_q = 0.31317.$$

Bond prices have been converted to the yield curve through

$$\bar{r}^*(0{:}k) = P^*(0{:}k)^{-1/k} - 1,$$

which is the average rate of interest over the period in question; see Section 1.4. The initial rate r_0 varied between $r_0 = 2\%$, 4% 6% 8% and 10% which produced the different shapes in Figure 6.2. In the long run the average yield tends to $\xi = 4\%$ with a speed determined by a_q.

1.5 Stochastic dependence: General

Introduction

General probabilitic descriptions of dependent random variables $X_1, \ldots, X_n$ are provided by **joint density functions** $f(x_1, \ldots, x_n)$ or **joint distribution functions** $F(x_1, \ldots, x_n)$. The latter are defined as the probabilities

$$F(x_1, \ldots, x_n) = \Pr(X_1 \leq x_1, \ldots X_n \leq x_n).$$

Its n-fold partial derivative with respect to $x_1, \ldots, x_n$ is (when its exists) the density function $f(x_1, \ldots, x_n)$ which may also be interpreted as the likelihood of the event

$$X_1 = x_1,\ X_2 = x_2,\ \ldots,\ X_n = x_n,$$

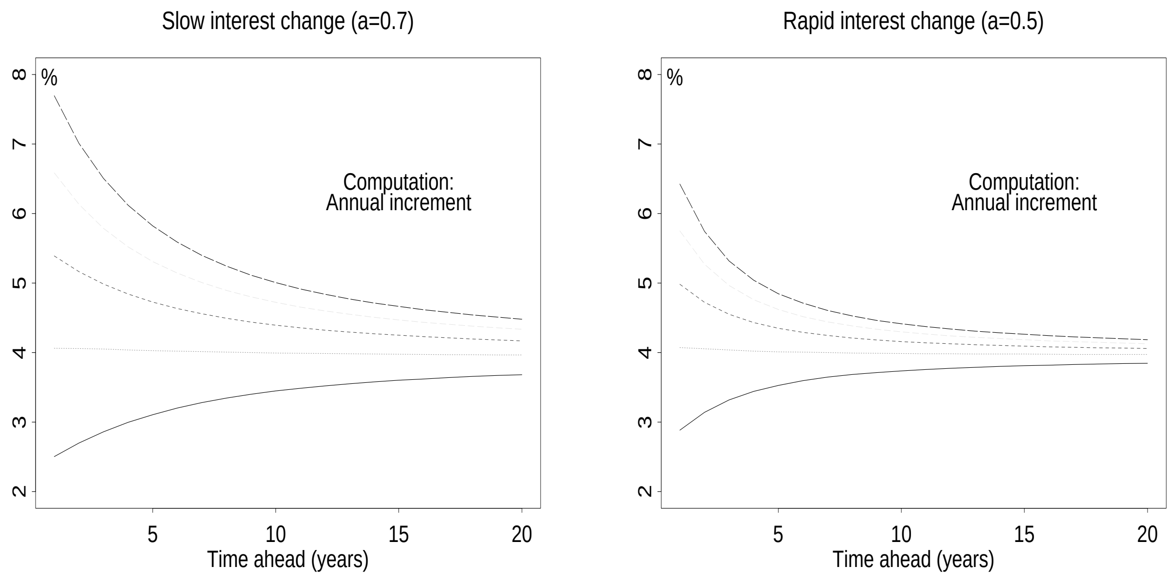


Figure 6.2 *Interest rate curves (from* $m = 10000$ *simulations) under the Black-Karisinsky model when the initial rate of interest is varied.*

though formally (in a strict mathematical sense) such probabilities are zero for continuous variables. Textbooks in probability and statistics often *start* with density functions. We need them for parameter estimation in the next chapter, but they are otherwise not much used in this book (with more advanced stochastic modelling they may play a vital role in checking logical consistency, but that is with our problems always obvious). Copulas in Section 6.7 are examples of modelling joint density functions directly.

Factorization of density functions

Whether $X_1, \ldots, X_n$ is a series in time or not we may always envisage them in a certain order. This observation opens for general simulation technique. Simply go recursively through the scheme

Sample	X_1^*	$X_2^*\|X_1^*$	$\cdots$	$X_n^*\|X_1^*, \ldots, X_{n-1}^*$
Probabiltities	$f(x_1)$	$f(x_2\|X_1^*)$	$\cdots$	$f(x_n\|X_1^*, \ldots, X_{n-1}^*)$,

where each drawing is conditional on what has come up before. We start by generating X_1 and end with X_n given all the others. The order selected does not matter in theory, but in practice there is often a natural sequence to use. If it isn't, look for other ways to do it.

The sampling scheme reflects the general factorization of joint density functions. Multiply the conditional ones together, and you get

$$\underbrace{f(x_1, \ldots, x_n) = f(x_1)f(x_2|x_1)\cdots f(x_n|x_1, \ldots, x_{n-1}).}_{\textit{general factorization}} \tag{1.29}$$

Note the notation where all density functions are denoted f though they differ according to their arguments. A special case is **Bayes'** formula. One version is

$$f(x_1|x_2, \ldots, x_n) = \frac{f(x_2, \ldots, x_n|x_1)f(x_1)}{f(x_2, \ldots, x_n)} \tag{1.30}$$

where the conditional density function of X_1 given $x_2, \ldots, x_n$ is referred back to the opposite form of $X_2, \ldots, X_n$ given x_1. This type of identity is crucial for Bayesian estimation in Section 7.6.

Types of dependence
Several special cases of (1.29) are of interest. The model with a common random factor in Section 6.2 is of the form

$$\underbrace{f(x_1, \ldots, x_n) = f(x_1)f(x_2|x_1)\cdots f(x_n|x_1)}_{\textit{Common factor: First variable}}. \tag{1.31}$$

Here the conditional densities only depend on the first variable, and all the variables $X_2, \ldots, X_n$ are conditionally independent given the first. Full independence means

$$\underbrace{f(x_1, \ldots, x_n) = f(x_1)f(x_2)\cdots f(x_n)}_{\textit{Independence}}. \tag{1.32}$$

Finally, there is the issue of **Markov dependence**, typically associated with time series. If X_k is attached to time t_k for $k = 1, \ldots n$, the model is now

$$\underbrace{f(x_1, \ldots, x_n) = f(x_1)f(x_2|x_1)\cdots f(x_n|x_{n-1})}_{\textit{Markov dependence}}, \tag{1.33}$$

where X_k only depends on the preceding x_{k-1} with the earlier $x_{k-2}, x_{k-3}, \ldots$ being irrelevant. Most models in life insurance belongs to this class, and the random walk and first order autoregression models of Section 5.6 do too; see below. How the general sampling scheme above is adapted is obvious, but the Markov situation is so important that the steps have been summarized through the following algorithm:

Algorithm 6.2 Markov sampling
0 Input: Conditional models
1 Generate X_1^*
2 For k=2,...,n do
3 Generate X_k^* given X_{k-1}^* *%Sampling from* $f(x_k|X_{k-1}^*)$

4 Return $X_1^*, \ldots, X_n^*$

Examples are given in Section 6.6 and in Exercise 6.5.1.

Linear and normal processes
All the time series models of Chapter 5 could have been introduced as Markov processes through a sequence of conditional distributions. As an example, consider the Vasicĕk model (**??**) which reads

$$r_k = r_{k-1} + (1-a)(\xi - r_{k-1}) + \sigma\varepsilon_k = \xi + a(r_{k-1} - \xi) + \sigma\varepsilon_k$$

where $\varepsilon_1, \varepsilon_2, \ldots$ are independent with zero mean and unit variance. Suppose they are normal too. Then

$$r_k|r_{k-1} = r \qquad \text{has the distribution of} \qquad \xi + a(r-\xi) + \sigma\varepsilon_k, \;\; \varepsilon_k \sim N(0,1),$$

and a model for the series $\{r_k\}$ is constructed by iterating over $k = 1, 2 \ldots$, first specify $r_1|r_0$, then $r_2|r_1$ and so on. This brings no particular benefit over the approach in Chapter 5, and indeed, the dynamic properties of the model were in Section 5.6 derived without introducing the normal.

The sequence $r_1, \ldots, r_k$ inherits Gaussianity from the errors $\varepsilon_1, \ldots, \varepsilon_k$, and it *is* possible to write down the joint density function. In the general case where $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_n)'$ is the vector of expectations and Σ the covariance matrix the Gaussian density function reads

$$f(\mathbf{x}) = (|2\pi\Sigma|)^{-1/2} \exp\{-\frac{1}{2}(\mathbf{x} - \boldsymbol{\xi})'\Sigma^{-1}(\mathbf{x} - \boldsymbol{\xi})\} \tag{1.34}$$

where $|2\pi\Sigma|$ is the determinant of the matrix $2\pi\Sigma$. This expression, though famous, plays no role in this book.

The multinomial situation

One joint density function that *will* be used is the multinomial one; see Section 8.5. This model is based on multinomial sampling (Section 4.2) where one label among K is selected according to probabilities $p_1, \ldots, p_K$ (here $p_1 + \ldots + p_K = 1$). This is repeated n times with each trial being independent of all others. Let N_k is the number of times label k appears. The vector $(N_1, \ldots, N_K)$ is then **multinomially** distributed with density function

$$\Pr(N_1{=}n_1, \ldots, N_K{=}n_K) = \frac{n!}{n_1! \ldots n_K!}\, p_1^{n_1} \ldots p_K^{n_K} \tag{1.35}$$

where $n_1 + \ldots + n_K = n$. Take $K = 2$ and we are back to the ordinary binomial.

The model (1.35) can be justified through a conditional argument. You have to be familiar with the binomial distribution. Let $\text{bin}(n, p)$ be the binomial density function based on n trials and success probabiltiy p and suppose (for simplicity) that $K = 3$. Then

$$N_1 \sim \text{bin}(n, p_1), \qquad N_2|N_1 = n_1 \sim \text{bin}\left(n - n_1, \frac{p_2}{p_2 + p_3}\right), \qquad N_3 = n - n_1 - n_2,$$

where N_3 is fixed by the two others. What lies behind the distribution stated for N_2 in the middle? We know that N_1 has absorbed n_1 trials. There are then $n - n_1$ of them left for N_2 and N_3 and among those label 2 must occur with probability $p_2/(p_2 + p_3)$.

The joint density function of (N_1, N_2) now becomes

$$\frac{n!}{n_1!(n - n_1)!}\, p_1^{n_1}(1 - p_1)^{n - n_1} \times \frac{(n - n_1)!}{n_2!(n - n_1 - n_2)!} \left(\frac{p_2}{p_2 + p_3}\right)^{n_2} \left(1 - \frac{p_2}{p_2 + p_2}\right)^{n - n_1 - n_2},$$

and if this is multplied out, you will discover that many of the factors cancel. The expression simplifies to

$$\frac{n!}{n_1! n_2! (n - n_1 - n_2)!}\, p_1^{n_1} p_2^{n_2} (1 - p_1 - p_2)^{n - n_1 - n_2} = \frac{n!}{n_1! n_2! n_3!}\, p_1^{n_1} p_2^{n_2} p_3^{n_3}$$

which is (1.35) when $K = 3$.

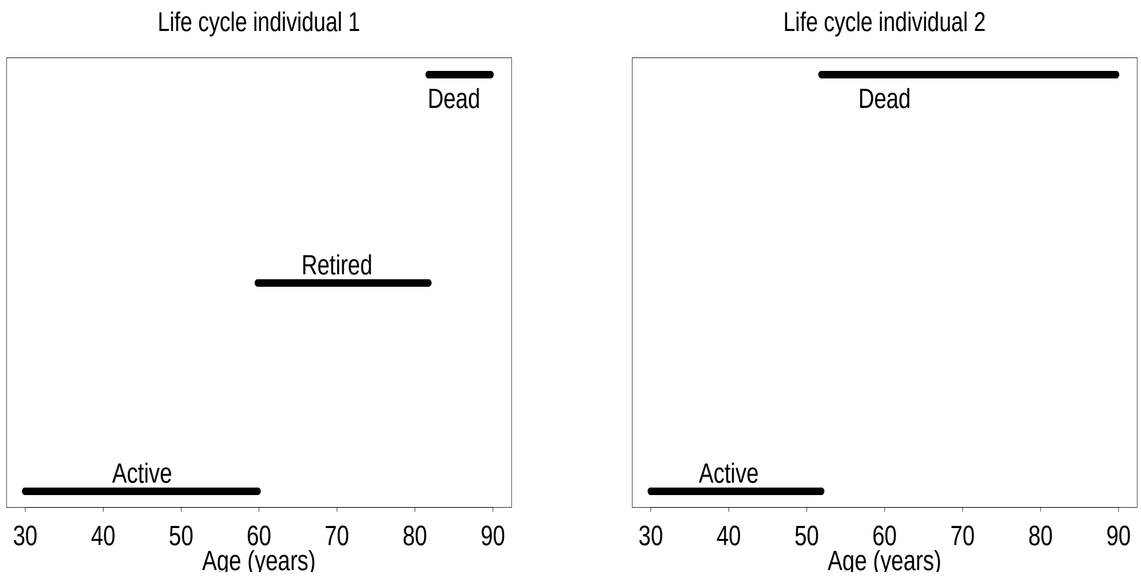


Figure 6.3 *The life cycles of two members of a pension scheme.*

1.6 Markov chains and life insurance

Introduction

Liability risk in life, pension and disability insurance are based on probabilistic descriptions of life cycles like those in Figure 6.3. The individual on the left dies at 82 having retired 22 years earlier at 60, whereas the other is a premature death at 52. A pension scheme consists many (even millions!) of members like those, each with his individual life cycle. Disability is a little more complicated, since there might be transitions back and forth; see below. It is worth noting that a switch from active to retired is determined by a clause in the contract, whereas death and disability must be described in random terms.

Each of the categories of Figure 6.3 will be called a **state**. A life cycle is a sequence $\{C_l\}$ of such states with C_l being the category occupied by the individual at age $y_l = lh$. We may envisage $\{C_l\}$ as a step function, jumping occasionally from one state to another. There are three of them in Figure 6.3. This section demonstrates how such schemes are described mathematically. Is it really required? After all, we saw in Section 3.4 that uncertainty due to life cycle movements is in insurance rarely very important. However, that does not make the underlying stochastic model irrelevant. It is needed both to compute the expectations defining the liabiltities (Chapter 12) *and* to evaluate portfolio uncertainty due to parameter error (Section 15.2).

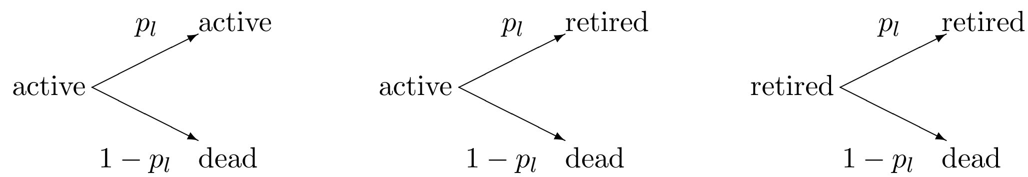
Markov modelling

Consider random step functions $\{C_l\}$ jumping between a limited number of states. The most frequently applied model is the **Markov chain**. What makes such time series evolve is the so-called **transition probabilities**

$$p_l(i|j) = \Pr(C_{l+1} = i|C_l = j). \tag{1.36}$$

At each point in time there is a random experiment that takes the state from its current j to a (possibly) new i. Note that the probabilitites defining the model *do not depend on the track record* of the individual prior to age l. That is the Markov assumption. Monte Carlo is a good way to find out how such models work; see Exercises 6.6.2 and 6.6.5.

Transition probabilities are usually different for men and women (not reflected in the mathematical notation), and it is (of course) essential that they depend on age l. There is always a main role for the survival probabilities $p_l = {}_1p_l$ introduced in Section 6.2. For a simple pension scheme, such as in Figure 6.3, the three states 'active', 'retired' and 'dead' are linked with the transition probabilities shown.

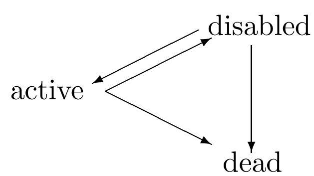


Before retirement *At retirement* *After retirement*

The details differ according to whether we are before, at or after retirement. Note the middle diagram in particular, where the individual from a clause in the contract moves from active to retired (unless he dies).

A disability scheme

Disability modelling, with movements back and forth between states, is more complicated. Consider the following scheme.



where

$$p_{i|a} = \Pr(\text{disabled}|\text{active})$$

$$p_{a|i} = \Pr(\text{active}|\text{disabled}).$$

A person may become 'disabled' (state i), but there is also a chance that he returns to 'active' (state a). Such rehabilitations are not too frequent as this book is being written (2008), but it could be different in the future, and we should certainly be able to to handle it mathematically. New probabilities are then needed in addition to those describing survival. They have above been denoted $p_{i|a}$ and $p_{a|i}$. Both will in practice depend on age l, but this aspect is here omitted. The probability of moving from 'active' to 'disabled' is denoted $p_{i|a}$ while the opposite is $p_{a|i}$.

The transition probabilities for the scheme must combine survival and disability/rehabilitation. The full matrix are as follows:

	To new state			
From	Active	Disabled	Dead	Row sum
Active	$p_l \cdot (1 - p_{i\|a})$	$p_l \cdot p_{i\|a}$	$1 - p_l$	1
Disabled	$p_l \cdot p_{a\|i}$	$p_l \cdot (1 - p_{a\|i})$	$1 - p_l$	1
Dead	0	0	1	1

Each entry is the product of input probabilities. For example, to remain active (upper left corner) the indvidual must survive *and* not become disabled, and similar for the others. Note the row sums. *They are always equal to one* (add them and you see it is true). *Any* set of transition probabilities for Markov chains must satisfy this restriction, which merely reflects that the individual either moves somewhere or remains where he is.

Numerical example

Figure 6.4 shows a portfolio development that might occur in practice. The survival model was the same as in Section 3.4, i.e.

$$\log(p_l) = -0.0009 - 0.0000462 \exp(0.090767 \times l)$$

Their corresponding annual **mortalitites** $q_l = 1 - p_l$ are plotted in Figure 6.4 left. Note the steep increase on the right for the higher age groups where the likelihood of dying within the coming year has reached 1.5% and more.

This model corresponds to an average length of life of 75 years and will be further discussed in Chapter 12. It is reasonably realistic for males in a developed country. Disability depends on the current political climate and on economic cycles and is harder to hang numbers on. The computations in Figure 6.4 are based on

$$p_{i|a} = 0.7\%, \qquad \text{and} \qquad p_{a|i} = 0.1\%,$$

which are values invented. Note the rehabilitation rate, which may be too high. In practice both probabilities might depend on age, as noted above.

How individuals distribute between the three states are shown in Figure 6.4 right for a portfolio originally consisting of one million 30-year males. The scenario has been simulated using Algorithm 6.2 (details in Exercise 6.6.2). There is very little Monte Carlo uncertainty in portfolios this size and one *single* run is enough. At the start all are active, but with age the number of people in the other two classes grow. At 65 years a little over 80% remain alive (distributed between 'active' and 'disabled'), a realistic figure. What may not be true in practice is the downwards curvature in the disability curve which might be turned around if the disability rate is made age-dependent.

1.7 Introducing copulas

Introduction

Let U_1 and U_2 be uniform, *dependent* random variables and consider

$$X_1 = F_1^{-1}(U_1) \qquad \text{and} \qquad X_2 = F_2^{-1}(U_2), \tag{1.37}$$

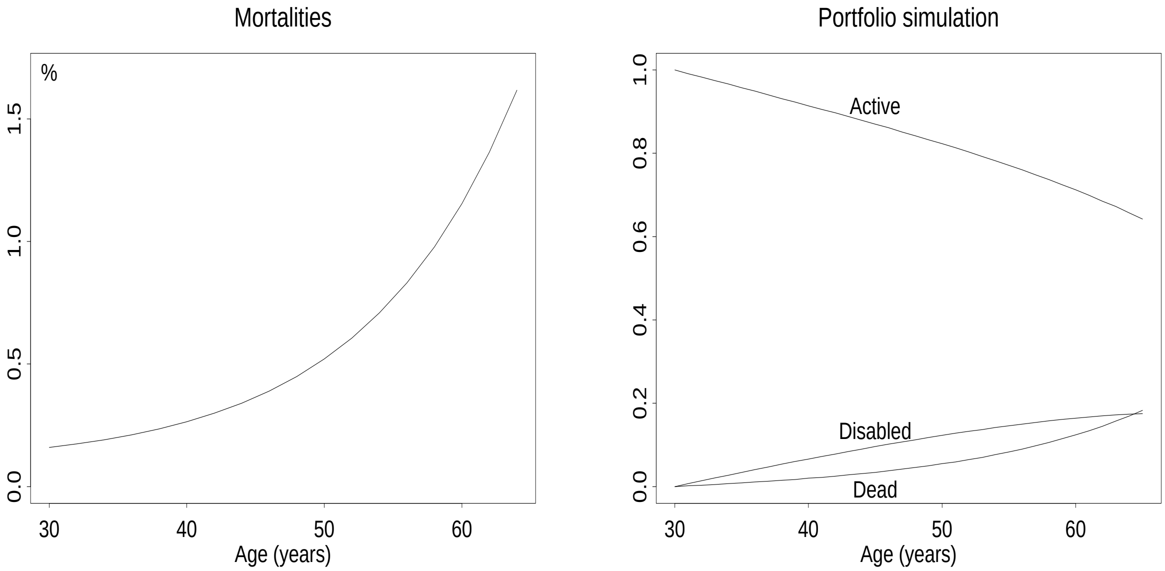


Figure 6.4 *A disability sceme in life insurance: Mortality model (left) and portfolio simulation (right).*

where $F_1^{-1}(u)$ and $F_2^{-1}(u)$ are the percentiles of distribution functions $F_1(x)$ and $F_2(x)$. This simple set-up defines a modelling strategy that has grown explosively popular. Treat dependence and univariate variation as two separate issues and supply a joint distribution function $C(u_1, u_2)$, known as a **copula**, to describe the covariation between U_1 and U_2 which is then passed on to X_1 and X_2. The inversion algorithm of Chapter 2 tells us that the marginal distributions of X_1 and X_2 are determined by the percentiles used in (1.37).

The idea goes back to the mid twentieth century, originating with the work of Sklar (1959). One of the advantages is that it enables us to tackle situations as those in Figure 6.5 where the variables are more strongly related in certain parts of the space than in others. An example in practice is equities that may correlate stronger in falling markets than on average; see Longin and Solnik (2001) for a formal, statistical justification. Such phenomena is of interest in insurance too. Wütrich (2004) is a theoretical contribution.

Copula modelling

A bivariate copula is the joint distribution function

$$C(u_1, u_2) = \Pr(U_1 \leq u_1, U_2 \leq u_2), \qquad 0 < u_1 \leq 1,\ 0 < u_2 \leq 1, \tag{1.38}$$

where $0 < u_1 \leq 1$ and $0 < u_2 \leq 1$. Any function $C(u_1, u_2)$ that is to play this role must be increasing in u_1 and u_2 and satisfy

$$C(u_1, 0) = 0,\ \ C(u_1, 1) = u_1 \qquad \text{and} \qquad C(0, u_2) = 0,\ \ C(1, u_2) = u_2. \tag{1.39}$$

Simple examples are

$$\underset{\textit{independent copula}}{C(u_1, u_2) = u_1 u_2} \qquad \text{and} \qquad \underset{\textit{Clayton copula}}{C(u_1, u_2) = (u_1^{-\theta} + u_2^{-\theta} - 1)^{-1/\theta}}$$

where $\theta > 0$ on the right. Both satisfy the side conditions (1.39). Why must we impose the latter? Because $C(u_1, 0) = \Pr(U_1 \leq u_1, U_2 \leq 0) = 0$ and $C(u_1, 1) = \Pr(U_1 \leq u_1, U_2 \leq 1) = \Pr(U_1 \leq u_1) = u_1$, no uniforms without this! The rationale for the other conditions in (1.39) is similar.

The copula approach rests on a representation theorem discovered by Sklar (1959). Any joint distribution function $F(x_1, x_2)$ with strictly increasing marginal distribution functions $F_1(x_1)$ and $F_2(x_2)$ may be written

$$\underbrace{F(x_1, x_2) = C(u_1, u_2)}_{\textit{copula modelling}} \qquad \text{where} \qquad \underbrace{u_1 = F_1(x_1), \;\; u_2 = F_2(x_2).}_{\textit{univariate modelling}} \tag{1.40}$$

with a modified version even for counts. Working through copulas do not restrict the model at all, and there are additional versions when antitetic twins (Section 4.3) are supplied for either uniform. The copula on the left in (1.40) may be combined with either of

$$\begin{array}{ll} u_1 = F_1(x_1), \;\; 1 - u_2 = F_2(x_2) & \textit{orientation (1,2)} \\ 1 - u_1 = F_1(x_1), \;\; u_2 = F_2(x_2) & \textit{orientation (2,1)} \\ 1 - u_1 = F_1(x_1), \;\; 1 - u_2 = F_2(x_2) & \textit{orientation (2,2),} \end{array} \tag{1.41}$$

and the effect is to rotate the copula patterns 90°, 180° and 270° compared to the orginal one (orientation $(1, 1)$); see Figure 6.5.

Extension to many variables is straightforward. A J-dimensional copula $C(u_1, \ldots, u_J)$ is the distribution function of J dependent uniform variables $U_1, \ldots, U_J$ and satisfies consistency requirements similar to those in (1.39); see Exercise ??. Transformations back to the original variables are now through $X_1 = F_1^{-1}(U_1), \ldots, X_J = F_1^{-1}(U_J)$, and there are 2^J ways of rotating patterns through the use of antitetic twins, not just 4. You calculate copulas of sub-vectors by inserting ones, for example the (joint) distribution function of $U_1, \ldots, U_j$ is $C(u_1, \ldots, u_j, 1, \ldots, 1)$.

Example: The Clayton copula

The Clayton copula was introduced above above, but its defintion can be extended to

$$C(u_1, u_2) = \max\left(u_1^{-\theta} + u_2^{-\theta} - 1)^{-1/\theta}, 0\right) \tag{1.42}$$

and it is easy to check that the copula requirements (1.39) are satisfied whenever $\theta \geq -1$. Nor is it a difficult exercise to show that $C(u_1, u_2) \to u_1 u_2$ as $\theta \to 0$, and $\theta = 0$ corresponds to U_1 and U_2 being independent. When θ is negative,

$$C(u_1, u_2) = 0 \qquad \text{if} \qquad u_2 > (1 - u_1^{-\theta})^{-1/\theta},$$

and certain pairs (u_1, u_2) are forbidden territory; see also Figure 6.6 below. Hard restrictions of that kind are often undesirable. Yet when negative θ is included, the family in a sense cover the entire range of dependency that is logically possible.; see Exercises 6.7.2.and 6.7.3.

Simulated structures generated by the Clayton copula are shown in Figure 6.5. The marginal distributions of X_1 and X_2 were normal with mean $\xi = 0.005$ and volatility $\sigma = 0.05$, precisely as in Figure 2.5 (and realistic for equity returns). Most striking is the cone-shapes patterns which

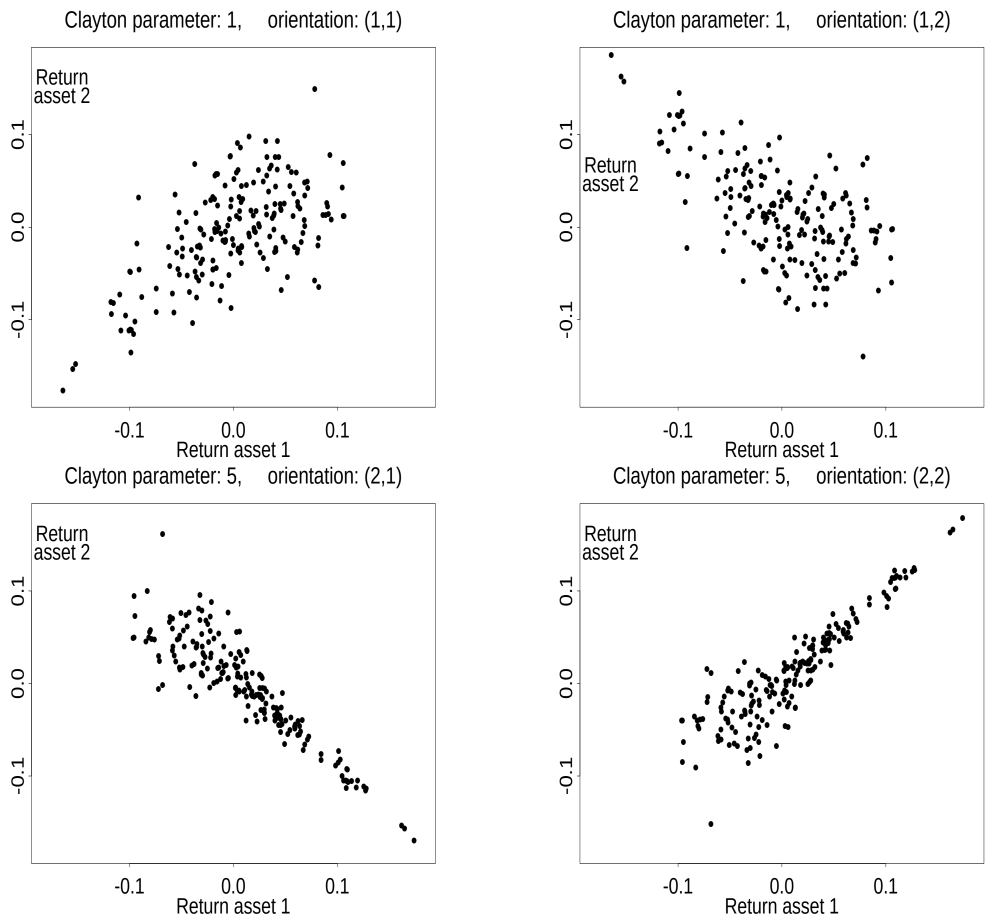


Figure 6.5 *Simulated financial returns from normals and* **Clayton** *copula.*

signify unequal degree of dependence in unequal parts of the space. Note, for example, the plot in the upper, left corner where downside returns are higher correlated than upside ones. Such situations have been seen in practice; consult Longin and Solnik (2001). Consequences for downside financial risk could be serious, and ordinary Gaussian models do not capture this. The other plots in Figure 6.5 rotate patterns by varying the orientation of the copula and adjust dependency by changing the Clayton parameter θ (high values for strong dependence).

Conditional distributions under copulas

Additional insight is provided by examining conditional distributions, and this is one of the ways copulas are simulated too. We shall need an expression for

$$C(u_J|u_1,\dots,u_{J-1}) = \Pr(U_J \le u_J|u_1,\dots,u_{J-1}).$$

In the two-dimensional case $J=2$

$$c(u_2|u_1) = c(u_1,u_2) = \frac{\partial^2 C(u_1,u_2)}{\partial u_1 u_2},$$

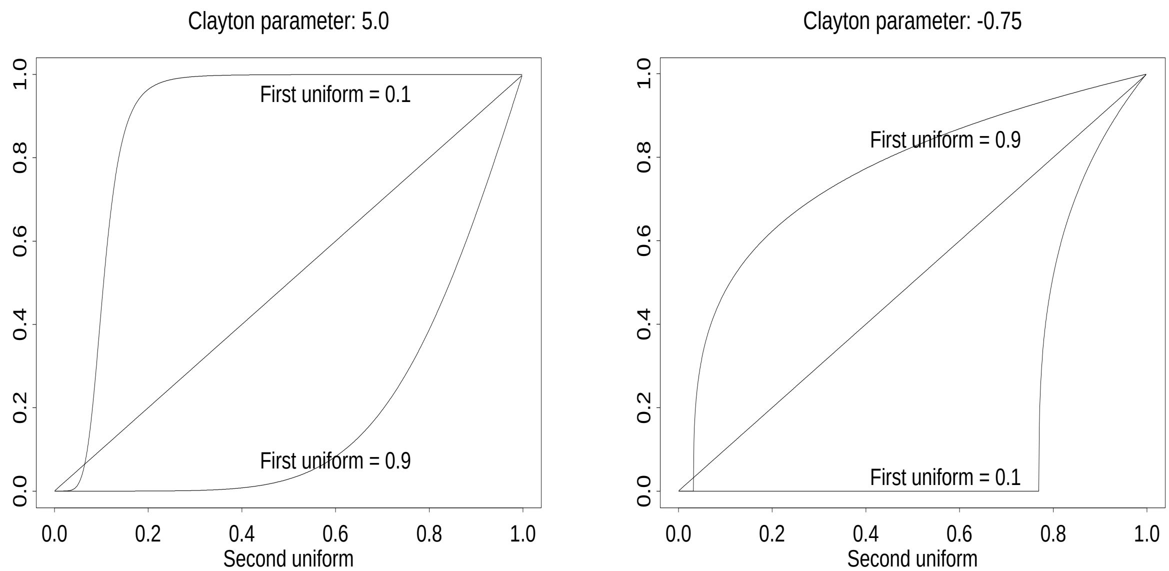


Figure 6.6 Conditional distribution functions for the **second** variable of a Clayton copula; given first variable marked on each curve.

where $c(u_1, u_2)$ is the joint density function of (U_1, U_2). When this is integrated with respect to u_2

$$C(u_2|u_1) = \int_0^{u_2} c(v|u_1)\,dv = \int_0^{u_2} \frac{\partial^2 C(u_1, v)}{\partial u_1 \partial v}\,dv = \frac{\partial}{\partial u_1}\int_0^{u_2} \frac{\partial C(u_1, v)}{\partial v}\,dv = \frac{\partial C(u_1,u_2)}{\partial u_1},$$

and conditional distribution functions are determined by differentiating the original copula. The general result is

$$C(u_J|u_1, \ldots, u_{J-1}) = \frac{\partial^{J-1}C(u_1, \ldots, u_{J-1}, u_J)/\partial u_1 \cdots \partial u_{J-1}}{\partial^{J-1}C(u_1, \ldots, u_{J-1}, 1)/\partial u_1 \cdots \partial u_{J-1}} \tag{1.43}$$

and you have to differentiate numerator and denominator $J-1$ times. This is proved in Section 6.8.

As an example, consider the Clayton copula (1.42) for $J = 2$. Straightforward calculations yields

$$C(u_2|u_1) = u_1^{-(1+\theta)} \max\left((u_1^{-\theta} + u_2^{-\theta} - 1)^{-(1+1/\theta)}, 0\right), \tag{1.44}$$

where the expression is zero when $\theta < 0$ and $u_2 > (1 - u_1^{-\theta})^{-1/\theta}$. Conditional distribution functions have been plotted in Figure 6.6 with θ is large and positive on the left and large and negative on the right. Shapes under $u_1 = 0.1$ and $u_2 = 0.9$ differ markedly, attesting to strong dependency between U_1 and U_2, but the most notable feature is a lack of symmetry. For the distributions on the left U_2 is located in a narrow strip around u_1 when $u_1 = 0.1$, but its variation is much larger when $u_1 = 0.9$. It is precisely this feature that creates the cones in Figure 6.5.

Archimedean copulas
These might be the most widely used of all copulas, and many of the most important ones have a

stochastic representation due to Marshall and Olkin (1988). Let Z be a positive random variable with density function $g(z)$. Its moment generating function (or Laplace transform) is

$$M(x) = E(e^{-xZ}) = \int_0^\infty e^{-xz} g(z)\,dz; \tag{1.45}$$

see Section A.1. Only positive x is of interest, and $M(x)$ decreases monotonely from one at $x = 0$ to zero at ∞. Define

$$U_j = M\left(-\frac{\log(V_j)}{Z}\right), \qquad j = 1, \ldots, J \tag{1.46}$$

where $V_1, \ldots, V_J$ is a sequence of independent and uniform random variables. Then $U_1, \ldots, U_J$ are uniform too (which isn't obvious!), and their joint distribution function is a copula of the form

$$C(u_1, \ldots, u_J) = \phi^{-1}\{\phi(u_1) + \ldots + \phi(u_J)\}, \tag{1.47}$$

where $\phi(u) = M^{-1}(u)$. Here $\phi^{-1}(u)$ and $M^{-1}(u)$ are the inverse functions of $\phi(u)$ and $M(u)$; for example $x = \phi^{-1}(u)$ is the solution of the equation $\phi(x) = u$. The result is proved in Section 6.8.

The Clayton coula emerges when Z is Gamma distributed with density function

$$g(z) = \frac{\alpha^\alpha}{\Gamma(\alpha)} z^{\alpha-1} e^{-\alpha z} \qquad \text{where} \qquad \alpha = 1/\theta.$$

Then

$$M(x) = \int_0^\infty e^{-xz} \frac{\alpha^\alpha}{\Gamma(\alpha)} z^{\alpha-1} e^{-\alpha z}\,dz = \left(1 + \frac{x}{\alpha}\right)^{-\alpha} = (1+\theta x)^{-1/\theta}$$

so that the two functions $\phi(u)$ and $\phi^{-1}(x)$ become

$$\phi(u) = \frac{1}{\theta}(u^{-\theta} - 1), \qquad \text{and} \qquad \phi^{-1}(x) = (1+\theta x)^{-1/\theta}. \tag{1.48}$$

When these are inserted into (1.47), the earlier expression for the Clayton copula emerges.

The function $\phi(u)$ is known as the **generator** and the copulas (1.47) as the **Archimedean** class. It was originally obtained by Kimberling (1974). Generators that are inverses of moment generating functions allow easy sampling (see below)[2], but generators are not confined to this type. Indeed suppose $\phi(u)$ decreases monotonely from infinity at $u = 0$ to zero at $u = 1$. Then, in the two-dimensional case,

$$C(u_1, 0) = \phi^{-1}\{\phi(u_1) + \phi(0)\} = \phi^{-1}(\infty) = 0$$

and

$$C(u_1, 1) = \phi^{-1}\{\phi(u_1) + \phi(1)\} = \phi^{-1}\{\phi(u_1)\} = u_1$$

which are the copula requirements (1.39). General J is similar. We may even allow finite $\phi(0)$, but the copula is now strictly zero whenever $\phi(u_1) + \ldots + \phi_J(u_J) > \phi(0)$, and certain combinations of $u_1, \ldots, u_J$ are forbidden. In practice this is rarely what we want. A huge list of generators are

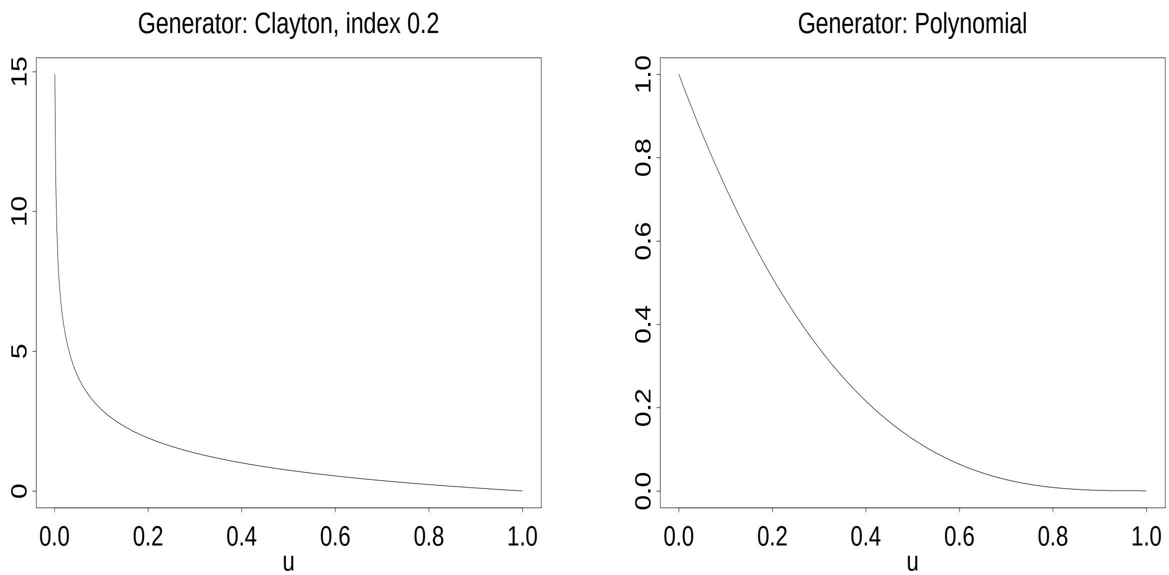


Figure 6.7 *Generator functions for Archimedean copulas*

compiled in Table 4.1 in Nelsen (2006). The Clayton generator with $\theta = 0.2$ and the polynomial one $\phi(u) = (1-u)^3$ are plotted in Figure 6.7.

Simulation
A Monte Carlo simulation $U_1^*, \ldots, U_J^*$ of a uniform vector under a copula is passed on to the original variables through

$$X_1^* \leftarrow F_1^{-1}(U_1^*), \ldots, X_J^* \leftarrow F_J^{-1}(U_J^*),$$

which is still another application of the inversion algorithm. There are simple exact or approximate ways of calculating percentiles for most distributions used in this book (Gamma and t-distributions are exceptions). But what about $U_1^*, \ldots, U_J^*$ itself? One class of models that *is* easy to handle are Archimedean copulas satisfying the Marshall-Olkin stochastic representation. Simply copy (1.46) as follows:

Algorithm 6.3 Archimedean copulas
0 Input: $\phi(u)$
1 Draw Z^* *%Z with Laplace transform* $M(u) = \phi^{-1}(u)$
2 For $j = 1, \ldots, J$ repeat
3   Draw $V^* \sim$ uniform   and $U_j^* \leftarrow -\log(V^*)/Z^*$
4 Return $U_1^*, \ldots, U_J^*$

When Z^* is drawn from the standard Gamma distribution with shape $\alpha = 1/\theta$, a simulation

[2]For the inverse $\phi(u)^{-1}(u)$ of a given generator to be a moment generating function of some distribution is must be **totally positive**; i.e its derivatives must satisfy $(-1)^s \dfrac{d^s\phi^{-1}(u)}{du^s} \geq 0$ for all s; see Feller (1971), p. 439.

of the Clayton copula is returned.

This doesn't work when $\theta < 0$, and other copulas are not in a form convenient for Monte Carlo at all. A general approach is to go recursively through the copula vector, first draw U_1^* (an ordinary uniform), then U_2^* given U_1^* and so on. Let $C(u_j|u_1,\dots,u_{j-1}) = \Pr(U_j \le u_j|u_1,\dots,u_{j-1})$ be the conditional distribution of U_j given the predecessors. Independent uniforms $V_1^*,\dots,V_J^*$ are passed on to a copula simulation by solving the equations

$$C(U_j^*|U_1^*,\dots,U_{j-1}^*) = V_j^*, \qquad j = 1,\dots,J. \tag{1.49}$$

Numerical methods are required in general. For Clayton copulas there is a neat algorithm:

Algorithm 6.4 The Clayton copula
0 Input: θ
1 Draw $U_1^* \sim$ uniform and $S^* \leftarrow 0$
2 For $j = 2,\dots,J$ do
3 $S^* \leftarrow S^* + (U_{j-1}^*)^{-\theta} - 1$ *%Updating from preceding uniform*
3 Draw $V^* \sim$ uniform
4 $U_j^* \leftarrow \{-S^* + (1+S^*)(V^*)^{-\theta/(\theta+j-1)}\}^{-1/\theta}$ *%Next uniform*
5 Return $U_1^*,\dots,U_J^*$

This algorithm, which has been used for copula simulations in this book is justified in Section 6.8.

Numerical example
Has copula modelling much practical impact? It depends (as always) on what the model is used for. Here is an illustration based on two financial assets with returns $R_1 = e^{\xi_1+\sigma_1\varepsilon_1}$ and $R_2 = e^{\xi_2+\sigma_2\varepsilon_2}$ where ε_1 and ε_2 are $N(0,1)$. Ordinary Gaussian modelling specifies the correlation $\rho = \text{cor}(\varepsilon_1,\varepsilon_2)$. Results under this model will now be compared with those under a Clayton copula. We are in the latter case assuming that $\varepsilon_1 = \Phi^{-1}(U_1)$ and $\varepsilon_2 = \Phi^{-1}(U_2)$ where $\Phi^{-1}(u)$ is the Gaussian percentile function. The models can be calibrated by taking $\rho = \text{cor}\{\Phi^{-1}(U_1),\Phi^{-1}(U_2)\}$ so that $\text{cor}(\varepsilon_1,\varepsilon_2)$ becomes the same under both models. A simple way is to use Monte Carlo and approximate ρ by

$$\rho^* = \frac{1}{m}\sum_1^m \Phi^{-1}(U_{1i}^*)\Phi^{-1}(U_{2i}^*)$$

where (U_{1i}^*, U_{2i}^*), $i = 1,,\dots,m$ are simulations.

The experiments reported in Figure 6.8 used the Calyton copula with $\theta = 1$ which corresponds to $\rho = 0.498$ in the Gaussian model. Monthly exectations and volatiltities $\xi_1 = \xi_2 = 0.005$ and $\sigma_1 = \sigma_2 = 0.05$ were the same as in Figure 6.5, and the model assumed corresponds to the scatterplot in the upper, right hand corner there. The density functions apply to the equally weighted portfolio $\mathcal{R} = R_1/2 + R_2/2$. Those on the left are for downside returns lower than $b = -5\%$; i.e. they are conditional densities of $\mathcal{R}$ given $\mathcal{R} < -5\%$, and have been computed by discarding all samples for which the portfolio return exceeds 5%. There are considerable discrepancies under the two models, but those go away for five-year returns on the right. Now the differences can hardly

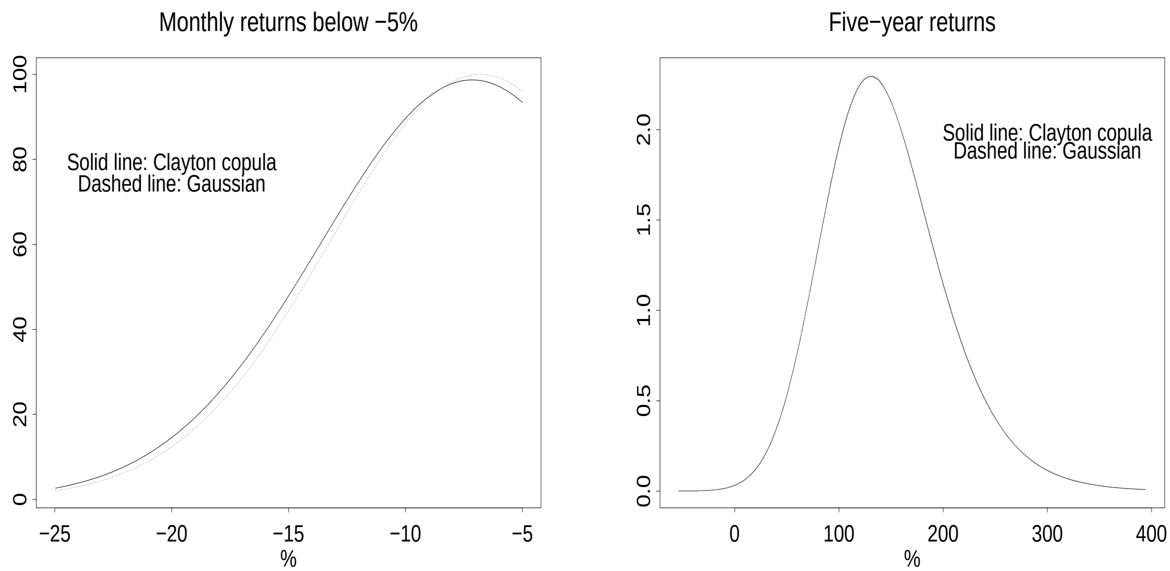


Figure 6.8 *Density functions for the monthly returns below* 5% *(***left***) and five-year returns (***right***) under models described i the text.*

be seen (though there are two curves if you look hard). Five-year (or sixty-month) returns are computed through the recursion

$$\mathcal{R}^*_{0:k} = \mathcal{R}^*_{0:k-1}(1+\mathcal{R}^*_k), \qquad k = 1,\dots,60$$

where $\mathcal{R}^*_{0:0} = 100$ and $\mathcal{R}^*_1,\dots,\mathcal{R}^*_{60}$ are independent drawings of the portfolio returns. Copulas effects as those under the Clayton model for $\theta = 1$ can evidently be ignored for long-range projections of financial risk.

1.8 Mathematical arguments

Section 6.3

Portfolio risk We shall verify the formula (1.22) for the standard deviation of the portfolio risk $\mathcal{X}$. Start with

$$\mathrm{var}(\mathcal{X}) = \mathrm{var}(\mathcal{N})\xi_z^2 + E(\mathcal{N})\sigma_z^2,$$

which is the right hand side of (1.19). Here $E(\mathcal{N}) = J\xi_\mu T$ whereas the variance varied with the sampling regime. Let $\gamma = J$ for common and $\gamma = 1$ for independent sampling of the claim intensity. Then the standard deviation formulae in (1.17) and (1.18) may be summarized as

$$\mathrm{var}(\mathcal{N}) = JT^2(\gamma\sigma_\mu^2 + \xi_\mu/T).$$

Inserting the expressions for $E(\mathcal{N})$ and $\mathrm{var}(\mathcal{N})$ into the formula for $\mathrm{var}(\mathcal{X})$ yields

$$\mathrm{var}(\mathcal{X}) = JT^2(\gamma\sigma_\mu^2 + J\xi_\mu/T)\xi_z^2 + J\xi_\mu T\sigma_z^2 = JT\xi_\mu(\sigma_z^2+\xi_z^2) + JT^2\gamma\sigma_\mu^2.$$

or

$$\operatorname{var}(\mathcal{X}) = \Big(JT\xi_\mu(\sigma_z^2+\xi_z^2)\Big) \;\times\; \left(1+\gamma T\frac{\sigma_\mu^2}{\xi_\mu}\frac{\xi_z^2}{\xi_z^2+\sigma_z^2}\right),$$

which is (1.22).

Section 6.4.
Optimal prediction To prove that $\hat{Y} = E(Y|\mathbf{X})$ is the optimal predictor for Y start by noting that

$$E(Y-a)^2 = E(Y^2) - 2aE(Y) + a^2, \qquad \text{minimized by} \qquad a = E(Y),$$

and apply this conditionally given $\mathbf{X}$. Then, if $\tilde{Y} = \tilde{Y}(\mathbf{X})$ is an arbitrary function of $\mathbf{X}$,

$$E\{(Y-\hat{Y})^2|\mathbf{X}\} \le E\{(Y-\tilde{Y})^2|\mathbf{X}\},$$

and when expectation with respect to $\mathbf{X}$ is passed over both sides, the rule of double expectation yields

$$E(Y-\hat{Y})^2 \le E(Y-\tilde{Y})^2$$

which is (1.25).

Section 6.7.
Conditional distributions We shall derive an expression for

$$C(u_J|u_1,\dots,u_{J-1}) = \Pr(U_J \le u_J|u_1,\dots,u_{J-1}) = \int_0^{u_J} c(v|u_1,\dots,u_{J-1})\,dv$$

where $c(u_J|u_1,\dots,u_{J-1})$ is the conditional distribution function of U_J given the others. To calculate the integral introduce $c(u_1,\dots,u_J)$ and $c(u_1,\dots,u_{J-1})$ as joint density functions of $U_1,\dots,U_J$ and $U_1,\dots,U_{J-1}$ and note that

$$c(u_J|u_1,\dots,u_{J-1}) = \frac{c(u_1,\dots,,u_J)}{c(u_1,\dots,u_{J-1})} = \frac{\partial^J C(u_1,\dots,u_J)/\partial u_1,\dots,\partial u_J}{\partial^{J-1}C(u_1,\dots,u_{J-1},1)/\partial u_1,\dots,\partial u_{J-1}}.$$

If D is the denominator, then

$$\begin{aligned}\int_0^{u_J} c(v|u_1,\dots,u_{J-1})\,dv &= D^{-1}\int_0^{u_J}\frac{\partial^J C(u_1,\dots,u_{J-1},v)}{\partial u_1,\dots,\partial u_{J-1},\partial v}\,dv\\ &= D^{-1}\frac{\partial^{J-1}}{\partial u_1,\dots,\partial u_{J-1}}\int_0^{u_J}\frac{\partial C(u_1,\dots,u_{J-1},v)}{\partial v}\,dv\\ &= D^{-1}\frac{\partial^{J-1}C(u_1,\dots,u_J)}{\partial u_1,\dots,\partial u_{J-1}},\end{aligned}$$

and it follows that

$$C(u_J|u_1,\dots,u_{J-1}) = \frac{\partial^{J-1}C(u_1,\dots,u_{J-1},u_J)/\partial u_1,\dots,\partial u_{J-1}}{\partial^{J-1}C(u_1,\dots,u_{J-1},1)/\partial u_1,\dots,\partial u_{J-1}}$$

as claimed in Section 6.7.

Stochastic representation of Archimedean copulas Let $V_1, \ldots, V_J$ be independent and uniform, Z a positive random variable with moment generating function $M(x) = \int_0^\infty e^{-xz} g(z) dz$ and define $U_j = M(-\log(V_j)/Z)$ for $j = 1, \ldots, J$. We are to prove that $U_1, \ldots, U_J$ follow an Archimedean copula with generator $\phi(u) = M^{-1}(u)$ provided Z is independent of $V_1, \ldots, V_J$. Note that $V_j = e^{-M^{-1}(U_j)Z}$ so that if $Z = z$ is fixed, then

$$\Pr(U_1 \leq u_1, \ldots, U_J \leq u_j | z) = \Pr(V_1 \leq e^{-M^{-1}(u_1)z} \ldots, V_J \leq e^{-M^{-1}(u_J)z} | z)$$

and since $V_1, \ldots, V_J$ are independent,

$$\Pr(U_1 \leq u_1, \ldots, U_J \leq u_j | z) = e^{-M^{-1}(u_1)z - \ldots - M^{-1}(u_J)z}.$$

But

$$\Pr(U_1 \leq u_1, \ldots, U_J \leq u_J) = \int_0^\infty \Pr(U_1 \leq u_1, \ldots, U_J \leq u_J | z) g(z)\, dz$$

and hence

$$\Pr(U_1 \leq u_1, \ldots, U_J \leq u_J) = \int_0^\infty e^{-\{M^{-1}(u_1) + \ldots + M^{-1}(u_J)\}z} g(z)\, dz$$

so that

$$\Pr(U_1 \leq u_1, \ldots, U_J \leq u_j) = M\{M^{-1}(u_1) + \ldots + M^{-1}(u_J)\}$$

which is a Archimedean copula with generator $\phi(u) = M^{-1}(u)$.

Algorithm 6.4 The Clayton copula for $U_1, \ldots, U_j$ is

$$C(u_1, \ldots, u_j) = \left(\max(\sum_{i=1}^{j} u_i^{-\theta} - j + 1, 0) \right)^{-1/\theta}.$$

which must be differentiated with respect to $u_1, \ldots, u_{j-1}$. This yields

$$\frac{\partial^{j-1} C(u_1, \ldots, u_j)}{\partial u_1 \ldots \partial u_{j-1}} = \left(\max(\sum_{i=1}^{j} u_i^{-\theta} - j + 1), 0) \right)^{-(1/\theta - j + 1)} \times \prod_{i=1}^{j-1} \left(u_i^{-(1+\theta)} (1 + (i-1)\theta) \right),$$

and from (1.43) the conditional distribution function of U_j given $u_1, \ldots, u_{j-1}$ becomes

$$\begin{aligned} C(u_j | u_1, \ldots, u_{j-1}) &= \left(\frac{\max(\sum_{i=1}^{j} u_i^{-\theta} - j + 1), 0)}{\max(\sum_{i=1}^{j-1} u_i^{-\theta} - j + 2), 0)} \right)^{1/\theta - j + 1} \\ &= \left(\frac{\max(u_j^{-\theta} + s_{j-1}), 0)}{\max(s_{j-1} + 1), 0)} \right)^{1/\theta - j + 1} \end{aligned}$$

where $s_{j-1}=\sum_{i=1}^{j-1}u_i^{-\theta}-(j-1)$. A Monte Carlo drawing of U_j^* given $U_1^*,\ldots,U_{j-1}^*$ is therefore the solution of the equation

$$\left(\frac{\max((U_j^*)^{-\theta}+S_{j-1}^*),0)}{\max(S_{j-1}^*+1),0)}\right)^{1/\theta-j+1}=V^* \qquad \text{where} \qquad S_{j-1}^*=\sum_{i=1}^{j-1}(U_i^*)^{-\theta}-(j-1)$$

and V^* is another uniform. It follows that

$$U_j^*=\{-S_{j-1}^*+(1+S_{j-1}^*)(V^*)^{-\theta/(\theta+j-1)}\}^{-1/\theta}$$

and the entire scheme can be organized as in Algorithm 6.4.

1.9 Bibliographical notes

Markov models and other basic topics in probability can be found in almost any textbook. The classic treatise Feller (1968) and (1971) is still a good place to start. Specialist monographs on multivariate distributions are Johnson, Kotz and Balakrishnan (1997) (the discrete type) and Kotz, Balakrishnan and Johnson (2000) (continuous ones); see also Balakrishnan (2004a) and (2004b) for the actuarial context on this. Among the applications the hierarchical risk model of Section 6.3 is a again a theme you will encounter in most textbooks on general insurance; see Mikosch (2004) for example. The Vasicĕk term structure in Section 6.4 is only one among a huge number of possibile models. There are countless others in James and Webber (2000), Briga and Mercurio (2001) and Cairns (2004). The interest in copulas simply exploded around the turn of the century. A skeptical eye is raised in Mikosch (2006). Good general introductions for insurance and finance are Embrechts, P, Lindskog and McNeil (2003) and Cherbini, Lucciano and Vecchiato (2004); see also Nelsen (2006) for a more mathematical treatment or even Joe (1997) for a broader angle on dependence modelling. You will find a good discussion of how Archimedean copulas are simulated in Whelan (2004); see also Frees and Valdez (1998). For applications in insurance you may consult (amomg others) Klugman and Parsa (1999), Carriére (2000), Venter (2003) and Escarela and Carriére (2007).

Balakrishnan, N. (2004a). Continuous Multivariate Distributions. In *Encyclopedia of Actuarial Science*, Teugels, J, and Sundt, B. (eds). John Wiley & Sons, Chichester, 549-570.

Balakrishnan, N. (2004b). Discrete Multivariate Distributions. In *Encyclopedia of Actuarial Science*, Teugels, J, and Sundt, B. (eds). John Wiley & Sons, Chichester, 357-362.

Briga, D. and Mercurio, F. (2001). *Interest Rate Models. Theory and Practice.* Springer, Berlin Heidelberg.

Cairns, A. (2004). *Interest Rate Models: an Introduction.* Princeton University Press, Princeton.

Carriére, J.F. (2000). Bivariate Survival Models for Coupled Lives. *Scandinavian Actuarial Journal*, 17-32.

Cherbini, U., Lucciano, E. and Vecchiato, W. (2004). *Copula Methods in Finance.* Johm Wiley & Sons, Chichester.

Embrechts, P., Lindskog, F. and McNeil, A. (2003). Modelling Dependence with Copulas and applications to risk management. In Rachev, T, ed. *Handbook of Heavy Tailed Distributions in Finance.* Elsvier, Amsterdam, 329-384.

Escarela, G. and Carriére, J. F. (2007). A Bivariate Model of Claim Frequencies and Severities.

Scandinavian Actuarial Journal, 8, 867-883.
Feller, W. (1968). *An Introduction to Probability Theory and its Applications.* Volume I. John Wiley & Sons, New York.
Feller, W. (1971). *An Introduction to Probability Theory and its Applications.* Volume II. John Wiley & Sons, New York
Frees, E.W. and Valdez, E. (1998). Understanding Relationships Using Copulas. *North American Actuarial Journal*, 2, 1-25.
James, J. and Webber, N. (2000). *Interest Rate Modelling.* John Wiley & Sons, Chichester.
Joe, H. (1997). *Multivariate Models and Dependence Concepts.* Chapman & Hall, London.
Johnson, N.L. Kotz, S. and Balakrishnan, N (1997). *Discrete, Multivariate Distributions.* John Wiley & Sons, New York.
Kimberling, C.H. (1974). A Probabilistic Interpretation of Complete Monotonicy. *Aequationes Mathematicae*, 10, 152-164.
Klugman, S.A. and Parsa, R. (1999). Fitting Bivariate Loss Distributions with Copulas. *Insurance: Mathematics and Economics*, 24, 139-148.
Kotz, S., Balakrishnan, N and Johnson, N.L. (2000). *Continuous, Multivariate Distributions.* John Wiley & Sons, New York.
Longin, F. and Solnik, B. (2001). Extreme Correlation of International Equity Markets. *Journal of Finance*, 56, 649-676.
Marshall, A. and Olkin, I. (1988). Families of Multivariate distributions. *Journal of American Statistical Association*, 83, 834-841.
Mikosch, T. (2004). *Non-Life Insurance Mathematics with Stochastic Processes.* Springer Verlag Berlin Heidelberg.
Mikosch, T. (2006). Copulas: Facts and Tales. *Extremes*, 9. 3-20.
Nelsen, R.B. (2006). Second ed. *An Introduction to Copulas.* Springer, New York.
Sklar, A. (1959). Fonctions de Répartion à n Dimensions et Leur Marges. *Publications de l'Institut de Statistique de l'Université de Paris*, 8, 229-231.
Vasicěk (1977). An Equilibrium Characterization of the Term Structure. *Journal of Financial Economics*, 5, 177-188.
Venter, C.G. (2003). Quatifying Correlated Reinsurance Exposures with Copulas. In Casualty Actuarial Society Forum, Spring, 215-229.
Whelan, N. (2004). Sampling from Archimedean Copulas. *Quantitative Finance*, 3, 339-352.
Wütrich, M.V. (2004). Extreme Value Theory and Archimedean Copulas. *Scandinavian Actuarial Journal*, 3, 211-228.

1.10 Exercises

Section 6.2

Exercise 6.2.1 The following experiment illustrates the concept of conditional distributions. Let $a_j = -0.5 + j/10$, for $j = 0, 1, \ldots, 10$. **a)** Simulate (X^*_{1i}, X^*_{2i}) for $i = 1, \ldots, 10000$ from the bivariate normal with $\xi_1 = \xi_2 = 5\%$, $\sigma_1 = \sigma_2 = 25\%$ and $\rho = 0.5$. **b)** For $j = 1, 2, \ldots, 9$, select those pairs for which $a_{j-1} < X^*_{1i} \leq a_j$ and compute their mean $\xi_{|j}$ and standard deviation $\sigma_{|j}$. **c)** Plot $\xi_{|j}$ and $\sigma_{|j}$ against the mid-points $(a_{j-1} + a_j)/2$, and interprete the plots in terms of the conditional density function (1.3). **d)** repeat a), b) and c) with $\rho = 0.9$ and comment on how the plot changes.

Exercise 6.2.2 Consider a time series $\{X_k\}$ of random variables such that the conditional distribution of X_k given all *preceding* ones are normal with

$$E(X_k|x_{k-1}, x_{k-2}, \ldots) = x_{k-1} + \xi \qquad \text{and} \qquad \text{sd}(X_k|x_{k-1}, x_{k-2}, \ldots) = \sigma.$$

Which of the times series models in Chapter 5 is this? see also Exercise 6.5.1.

Exercise 6.2.3 Let Z be a positive random variable and suppose X given $Z = z$ is normal with

$$E(X_|z)) = \xi \qquad \text{and} \qquad \text{sd}(X|z) = \sigma_0\sqrt{z}.$$

Which model from Chapter 2 is this?

Exercise 6.2.4 Let the survival probabiltities be those used in Section 3.4.; i.e.

$$\log({}_1p_l) = -0.0009 - 0.0000462\exp(0.090767 \times l).$$

a) For $l = 40$ and $l = 70$ years, compute ${}_kp_l$ as given in (1.6) and plot them as a function of k for $k = 1, 2, \ldots, 30$.

Exercise 6.2.5 Let N be an integer-valued random variable. **a)** Show that

$$\sum_{n=1}^{\infty} \Pr(N \geq n) = \sum_{n=1}^{\infty}\sum_{k=n}^{\infty} \Pr(N = k) = \sum_{k=1}^{\infty}\sum_{n=1}^{k} \Pr(N = k) = \sum_{k=1}^{\infty} k\Pr(N = k)$$

so that

$$E(N) = \sum_{n=1}^{\infty} \Pr(N \geq n).$$

Let N_l be the remaining length of life for somebody having reached l years. **b)** Use a) to establish that

$$E(N_l) = \sum_{k=1}^{\infty} {}_kp_l.$$

Exercise 6.2.6 Let X be an exponentially distributed random variable with density function $f(x) = \beta^{-1}\exp(-x/\beta)$ for $x > 0$. Show that in (1.8) $f_a(y) = f(y)$.

Exercise 6.2.7 Suppose that $f(x) = \beta^{-1}\alpha/(1 + x/\beta)^{1+\alpha}$ for $x > 0$ (this is the Pareto density). **a)** Show that

$$f_a(y) = \frac{\alpha/(a + \beta)}{\{1 + y/(a + \beta)\}^{1+\alpha}} \qquad \text{if} \qquad f_a(y) = \frac{\alpha/\beta}{(1 + y/\beta)^{1+\alpha}}$$

b) Interprete this result; i.,e what is the over-threshold distribution if the parent model is Pareto?

Exercise 6.2.8 A simple (but much less used) alternative to the gamma model to describe variation in the claim intensity μ is the log-normal. The model for portfolio claims then reads

$$\mathcal{N}|\mu \sim \text{Poisson}(J\mu T) \qquad \mu = \xi\exp(-\frac{1}{2}\sigma^2 + \sigma\varepsilon), \qquad \varepsilon \sim N(0, 1).$$

a) Show that

$$E(\mu) = \xi \qquad \text{and} \qquad \text{sd}(\mu) = \xi\sqrt{\exp(\sigma^2) - 1}$$

b) Determine σ so that $\text{sd}(\mu) = 0.1 \times \xi$. **c)** Run and plot simulations of $\mathcal{N}$ similar to those in Figure 6.2, using $\xi = 5\%$ and σ as you determined it in b). Take both $J = 10^4$ and $J = 10^6$ as portfolio size. **d)** Any conclusions that differ from those connected to Figure 6.2 in the text?

Section 6.3
Exercise 6.3.1 Suppose claim frequency $\mathcal{N} \sim \text{Poisson}(J\mu T)$. Show that the formulas (1.19) for mean and variance of the total claim $\mathcal{X}$ nowx become

$$E(\mathcal{X}) = J\mu T\xi_z \qquad \text{and} \qquad \text{sd}(\mathcal{X}) = \sqrt{J\mu T(\xi_z^2 + \sigma_z^2)}.$$

Exercise 6.3.2 Suppose claim intensities μ vary independently from one policy holder to another so that

$$\text{sd}(\mathcal{X}) = \sqrt{J\xi_\mu(\sigma_z^2 + \xi_z^2)} \times \sqrt{1+\delta} \qquad \text{where} \qquad \delta = \frac{\sigma_z^2}{\sigma_z^2 + \xi_z^2} \times \frac{\sigma_\mu^2}{\xi_\mu};$$

see (1.21) and (1.22). **a)** Show that $\delta \leq \sigma_\mu^2/\xi_\mu$. **b)** Argue that the case $\xi_\mu = 5\%$ and $\sigma_\mu = 5\%$ would exhibit huge variability in claim intensity. **c)** Use a) to show that $\sqrt{1+\delta} \leq 1.023$ under the specification in b) and argue that the *added* portfolio risk due to the hetereogenity in μ accounts for no more than 2.3% of the total value of $\text{sd}(\mathcal{X})$. This strongly suggests that at portfolio level the impact of risk hetereogenity usually can be ignored. The next exercise treats a related case where the conclusion is very different.

Exercise 6.3.3 As in Exercise 6.3.2 assume that μ varies randomly, but now as a common parameter for all policy holders. **a)** Go back to (1.21) and explain why the factor

$$\sqrt{1+J\delta} = \sqrt{1 + J\frac{\sigma_z^2}{\sigma_z^2 + \xi_z^2} \times \frac{\sigma_\mu^2}{\xi_\mu}};$$

accounts fot the effect of the *μ-variablity* on sd(§). **b)** Compute it when

$$\xi = 5\%, \qquad \sigma_\mu = 1\%, \qquad \frac{\sigma_z}{\xi_z} = 0.5 \qquad J = 100000.$$

Any comments? **c)** Show that the factor $\sqrt{1+J\delta}$ increases with the ratio σ_z/ξ_z. Is the impact of μ-variability larger or smaller for heavy-tailed claim size distributions than for lighter ones?

Exercise 6.3.4 Suppose that $X_1, \ldots, X_J$ are conditionally independent and identically distributed given a common factor ω. **a)** Explain that (1.15) now becomes

$$E(\mathcal{X}) = JE\{\xi(\omega)\} \qquad \text{and} \qquad \text{sd}(\mathcal{X}) = J\sqrt{\text{var}\{\xi(\omega)\} + E(\sigma^2(\omega)\}/J},$$

where $\xi(\omega)$ and $\text{sd}(\omega)$ are the conditional mean and standard deviation. **b)** Show that

$$\frac{\text{sd}(\mathcal{X})}{E(\mathcal{X})} \to \frac{\text{sd}\{\xi(\omega)\}}{E\{\xi(\omega)\}} \qquad \text{as} \qquad J \to \infty.$$

c) What this tell you about risk diversification models with common factors? This result throws light on the conclusion in Exercise 6.3.3

Exercise 6.3.5 Let $\mathcal{N}^*$ be a simulation of a Poisson claim frequency $\mathcal{N}$ where the intensity μ has been estimated as $\hat{\mu}$. If $T = 1$, this means that $\mathcal{N}^*|\hat{\mu}$ is $\text{Poisson}(J\hat{\mu})$. **a)** Use the double rules to prove that

$$E(\mathcal{N}^*) = JE(\hat{\mu}) \qquad \text{and} \qquad \text{var}(\mathcal{N}^*) = JE(\hat{\mu}) + J^2\text{var}(\hat{\mu}).$$

b) Recall that $E(\mathcal{N}) = \text{var}(\mathcal{N})$ for a Poisson variable $\mathcal{N}$ whereas $E(\mathcal{N}^*) < \text{var}(\mathcal{N}^*)$. What causes the difference? Integration of random error from different sources is discussed in Chapter 7.

Exercise 6.3.6 Suppose $X^* \sim N(\hat{\xi}, \hat{\sigma})$ where $\hat{\xi}$ and $\hat{\sigma}$ are estimates of ξ and σ from historical data. This should be interpreted as X^* having a conditional distribution given the estimates. **a)** Argue, using the double rules, that

$$E(X^*) = E(\hat{\xi}) \qquad \text{and} \qquad \text{var}(X^*) = E(\hat{\sigma^2}) + \text{var}(\hat{\xi})$$

b) Suppose that $\text{var}(\hat{\xi}) = \sigma^2/n$ and that $E(\hat{\sigma}^2) = \sigma^2$ (which you recognize as a standard situation with n historical observations). Show that

$$\text{var}(X^*) = \sigma^2(1 + \frac{1}{n}).$$

Exercise 6.3.7 The double rule for variances can be extended to a version for covariances. Let

$$\xi_1(\mathbf{x}) = E(Y_1|\mathbf{x}), \qquad \xi_2(\mathbf{x}) = E(Y_2|\mathbf{x}) \qquad \text{and} \qquad \sigma_{12}(\mathbf{x}) = \text{cov}(Y_1, Y_2)|\mathbf{x})$$

for random variables Y_1, Y_2 conditioned on $\mathbf{X} = \mathbf{x}$. Then

$$\text{cov}(Y_1, Y_2) = \text{cov}\{\xi_2(\mathbf{X}), \xi_1(\mathbf{X})\} + E\{\sigma_{12}(\mathbf{X})\};$$

see Appendix A. Use this to find the covariances bewteen returns R_1 and R_2 satisfying the stochastic volatility model in Section 2.4; i.e

$$R_1 = \xi_1 + \sigma_{01}\sqrt{Z}\varepsilon_1 \qquad \text{and} \qquad R_2 = \xi_2 + \sigma_{02}\sqrt{Z}\varepsilon_2$$

where ε_1, ε_2 and Z are independent and the two former are $N(0, 1)$ with correlation ρ.

Section 6.4

Exercise 6.4.1 Let X_1 and X_2 be dependent normal variables with expectations ξ_1 and ξ_2, standard deviations σ_1 and σ_2 and correlation ρ. **a)** Use (1.3) to justify that

$$\hat{X}_2 = \xi_2 + \rho\sigma_2\frac{x_1 - \xi_1}{\sigma_1} \qquad \text{for} \qquad X_1 = x_1$$

is the most accurate prediction of X_2 if X_1 is observed. **b)** Show that

$$\frac{\text{sd}(\hat{X}_2|x_1)}{\text{sd}(X_2)} = \sqrt{1 - \rho^2}.$$

c) By how much is the uncertainty in X_2 reduced by knowing X_1 if $\rho = 0.3$, 0.7 and 0.9? Argue that ρ should from this viewpoint be interpreted through ρ^2, as claimed in Section 5.2.

Exercise 6.4.2 Claim intensities μ in automobile insurance depends on factors such as age and sex. Consider a female driver of age x. A standard way to formulate the link between x and μ goes through the conditional mean $E(N|x)$, where N is claim frequency. One possibility is

$$\mu = \mu_0 e^{-\beta(x - x_0)},$$

where x_0 is the starting age for drivers and μ_0 and β_0 are parameters. **a)** What is the meraning of the parameters μ_0 and β? **b)** Determine them so that $\mu = 10\%$ at age 18 and 5% at age 60 and plot the relationship between x and μ. In practice a more complex relationship is often used; see Chapter 8.

Exercise 6.4.3 Let

$$\xi = 5\%, \qquad a = 0.5 \qquad \sigma = 0.016, \qquad r_0 = 2\%$$

in the Vasicěk model for interest rates. **a)** Write down predictions for the rate of interest r_k at $k = 1, 2, 5$ and $k = 10$ years, using (**??**). **b)** What is standard standard deviation of the prediction error? Use (**??**) and compare the assessment for $k = 1$ and $k = 5$ with those in Section 6.4 coming from a related (but different) set of parameters.

Exercise 6.4.4 Conside the Black-Karisisnski model defined in Section 5.7 under which

$$r_k = \xi \exp(-\frac{1}{2}\sigma_x^2 + X_k) \qquad \text{where} \qquad \sigma_x = \frac{\sigma}{\sqrt{1-a^2}}, \quad X_k = aX_{k-1} + \sigma\varepsilon_k.$$

Here $\varepsilon_1, \varepsilon_2, \ldots$ are all independent and $N(0, 1)$. **a)** If r_0 is the current rate of interest observed in the market, aregue that

$$\hat{r}_k = \xi \exp(-\frac{1}{2}\sigma_x^2 + a^k x_0) \qquad \text{where} \qquad x_0 = \log\left(\frac{r_0}{\xi} + \frac{1}{2}\sigma_x^2\right)$$

is a prediction of the future rate r_k. **b)** Make the prediction for $k = 1, 2, 5$ and $k = 10$ years as in the preceding exercise and use the same parameters as there. Compare forecasts under the two models. This example will be examined further in Exercise 7.?.

Exercise 6.4.5 Algorithm 6.1 dealt with the forward rate of interest under the Vasiceěk model. **a)** Modify it so that it applies to the Black-Karisisnksi model [Hint: You replace Line 3 with parts of Algorithm 5.4.]. **b)** ???

Section 6.5

Exercise 6.5.1 Suppose the time series $\{X_k\}$ is a Gaussian Markov process for which

$$X_k | X_{k-1} = x \;\sim\; N(ax, \sigma).$$

Which model from Chapter 5 is this?

Exercise 6.5.2 Suppose $X_1, \ldots, X_J$ are conditionally normal given $Z = z$ with expectations ξ_i and variance/covariances $\sigma_{ij} z$. **a)** Which model from earlier chapters is this? **b)** Do the *correlations* depend on z? Which model from Chapter 5 is this?

Exercise 6.5.3 Consider Algorithm 6.2, the skeleton for Markov sampling. **a)** Modify it to deal with *common factors*; i.e explain that X_k^* on Line 3 now is drawn from $f(x_k | X_1^*$.

Exercise 6.5.4 This exercise shows how a stochastic volatiltiy model for *log*-returns are sampled by means of the preceding exercise. Suppose

$$Z = \exp(-\frac{1}{2}\tau^2 + \tau\varepsilon), \qquad \varepsilon \sim N(0, 1)$$

is log-normal and that

$$X_1 = \log(1 + R_1), \quad X_2 = \log(1 + R_2), \quad X_3 = \log(1 + R_3)$$

are conditionally normal with expectations ξ_1, ξ_2, ξ_3, volatilities $\sigma_{01} sqrtz, \sigma_{02}\sqrt{z}, \sigma_{03}\sqrt{z}$ and correlations ρ_{ij}. **a)** Explain how the log-returns are samples. **b)** Carry out the sampling 1000 times when

$$\xi_1 = \xi_2 = \xi_3 = 5\%, \qquad \sigma_{01} = \sigma_{02} = \sigma_{03} = 0.2, \qquad \text{all } \rho_{ij} = 0.5 \quad \text{and} \qquad \tau = 0.5.$$

c) Use b) to compute the 5% lower percentile of the portfolio with equal weights on the three risky assets.

Exercise 6.5.5 Stochastic volatility in finance is in reality a *dynamic* phenomenon where the random variable $Z = Z_k$ being responsible are correlated in time. The first model proposed to deal with this is known as **ARCH**[3] and can be formulated as follows:

$$R_k = \xi + \sigma_0\sqrt{Z}_k\varepsilon_k \qquad \text{where} \qquad Z_k = \sqrt{1 + \theta(R_{k-1} - \xi)^2}$$

where $\varepsilon_1, \varepsilon_2, \ldots$ are independent and $N(0,1)$. **a)** Argue that returns deviating strongly from the mean ξ makes volatility go up next time. **b)** Why is this a Markov model for the series $\{R_k\}$? **c)** Simulate the model and plot the against time k for $k = 1, \ldots, 30$ when

$$\xi = 5\%, \qquad , \sigma_0 = 0.2 \quad \text{and} \quad \theta = 0.2 \qquad \text{starting at} \quad R_0 = 5\%.$$

These are annual parameters. Plot ten different scenarios.

Exercise 6.5.6 An alternative to ARCH of the preceding is to use the Black-Karisinski model from Section 5.7 for $\{Z_k\}$, i.e to take

$$Z_k = \exp(-\frac{1}{2}\tau_y^2 + \tau_y Y_k) \qquad \text{where} \qquad \tau_y = \frac{\tau}{\sqrt{1-a^2}}, \quad Y_k = aY_{k-1} + \tau\eta_k.$$

Here both sequences $\eta_1, \eta_2 \ldots$ and $\varepsilon_1, \varepsilon_2, \ldots$ are independent $N(0,1)$ and independent from each other. **a)** Simulate and plot ten realisations of this model under the same conditions as in the previous exercise using $a = 0.6$ and $\tau = 0.1$. **b)** Is there in behaviour a principal difference form the ARCH model. This model type, though less used than the former (and, especially its extensions) is drawing much interest as this book is being written (2004).

Exercise 6.5.7 The multinomial model illustrates the factorization (1.29). Start by noting that $N_0 \sim$Binomial(n, q_0). **a)** Then argue that

$$N_1|n_0 \sim \text{Binomial}(n - n_0, \tilde{q}_1) \qquad \text{where} \qquad \tilde{q}_1 = \frac{q_1}{1 - q_0}.$$

[Hint: From n trials originally, subtract those ($= n_0$) with no delay. Among the *remaining* $n - n_0$ trials the likelihood is $\tilde{q}_1$ for delay exactly one year.]. Suppose a binomual sampling procedure is available. **b)** Justify that (N_0, N_1) can be sampled through

$$N_0^* \sim \text{Binomial}(n, q_0) \qquad \text{and} \qquad N_1^* \sim \text{Binomial}(n - N_0^*, \tilde{q}_1)$$

The next step is

$$N_2^*|n_0, n_1 \sim \text{Binomial}(n - n_0 - n_1, \tilde{q}_2) \qquad \text{where} \qquad \tilde{q}_2) = \frac{q_2}{1 - q_0 - q_1}.$$

c) Explain why the general case can be run as follows:

Algorithm 6.6 Multinomial sampling

[3] ARCH stands for **a**uto**r**egressive, **c**onditional, **h**etereochedastic.

0 Input n and $q_0, \ldots, q_K$
1 $S^* \leftarrow 0, \quad d \leftarrow 1$
2 For $k = 1, \ldots, K-1$ do
3 Draw $N_k^* \sim$Binomial$(n - S^*, p_k^*/d)$
4 $S^* \leftarrow S^* + N_k^*, \quad d \leftarrow d - p_k$

5 Return $N_1^*, \ldots, N_{K-1}^*$ and $N_K^* \leftarrow n - S^*$.

This is inefficient for large K, but tolerable for delay. **d)** Run the algorithm 10000 times when

$$K = 4, \quad q_0 = 0.1, \quad q_1 = 0.3 \quad q_2 = 0.25, \quad q_3 = 0.2, \quad q_4 = 0.15.$$

and compare relative frequencies with the underlying probabiltites.

Exercise 6.5.8 We know from the preceding exercise that

$$\Pr(N_0 = n_0) = \frac{n!}{n_0!(n-n_0)!} q_0^{n_0}(1-q_0)^{n-n_0}$$

and that

$$\Pr(N_1 = n_1|n_0) = \frac{(n-n_0)!}{n_1!(n-n_0-n_1)!} \tilde{q}_1^{n_1}(1-\tilde{q}_1)^{n-n_0-n_1}.$$

Multiply the two probabilities together and verify that

$$\Pr(N_0 = n_0, N_1 = n_1) = \frac{n!}{n_0!n_1!(n-n_0-n_1)!} q_0^{n_0} q_1^{n_1}(1-q_0-q_1)^{n-n_0-n_1}.$$

This is the multinomial density function (1.35) for $K = 2$ (note that $N_2 = n - n_0 - n_1$ is fixed by the two first). The general case is established by continuing in this way.

Section 6.6

Exercise 6.6.1 Consider a Markov chain $\{C_k\}$ running over the three states "active", "disabled" and "dead" with $p^{a|d}$ and $p^{a|d}$ as probabilities of going from "disabled" to "active" and "active" to "disabled" and with probability of survival ${}_1p_l$ at age l. **a)** Argue, using conditioning, that the probability at age l of remaining active must be ${}_1p_l(1 - p^{a|d})$. **b)** Fill out the rest of the table of transition probabilities at page ?, using the same reasoning. **c)** Verify that the row sums are equal to one. **d)** What does the matrix become when

$$\log({}_1p_l) = -0.0009 - 0.0000462 \exp(0.090767 \times l), \qquad p^{d|a} = 0.7\%, \qquad p^{a|d} = 0.35\% \qquad ?$$

Exercise 6.6.2 Let the three states of the preceding exercise be labeled 0 (for "active"), 1 ("disabled") and 2 ("dead") and let $p_l(i|j)$ be their transition probabiltites at age l. **a)** Implement Algorithm 6.2 for the model of the preceding exercise. For example, argue that the following recursive step can be used on Line 3:

Draw $U^* \sim$ uniform and $l \leftarrow l + 1$
If $U^* < p_l(0|C_{k-1}^*)$ then $C_k^* \leftarrow 0$
else if $U^* < p_l(0|C_{k-1}^*) + p_l(2|C_{k-1}^*)$ then $C_k^* \leftarrow 1$
else $C_k^* \leftarrow 2$ and stop.

b) Run the algorithm ten times with the model of Exercise 6.1, each time starting at age $l = 30$ years and plotting the the simulated scenarios 50 years ahead. **c)** Change the model unrealistically!) to $p^{d|a} = 0.4$ and $p^{a|d} = 0.20$, re-compute the transitiom matrix and re-run the simualtions to see different patterns.

Exercise 6.6.3 The expected remaining life-time at age l was derived an Exercise 6.2.5 as

$$E(N_l) = \sum_{k=1}^{\infty} {}_kp_l \qquad \text{where} \qquad {}_kp_l = {}_1p_{l+k-1} \times \cdots \times {}_1p_l$$

Consider the recursion

$$P \leftarrow {}_1p_l \times P, \qquad E \leftarrow E + P, \qquad l \leftarrow l + 1$$

starting at $P = 1$, $E = 0$. **a)** Argue that it yields $E = E(N_l)$ at the end. **b)** Implement the recursion, compute $E(N_l)$ for $l = 20, 25, 30, \dots$ up to $l = 70$ for the survival model in Exercise 6.6.1. **c)** Plot the computed sequence against l and explain why it is decreasing.

Exercise 6.6.4 One of the issues with potentially huge impact on the business of life and pension insurance is the fact that in most countries length of life is steadily prolonged. Suppose we want to change our current survival model into a related one in order to get a rough picture of the economic consequences. A simple way is to introduce

$${}_1\tilde{p}_l = \frac{\theta {}_1p_l}{\theta {}_1p_l + (1 - {}_1p_l)},$$

where θ is a parameter. **a)** Show that the new survival probability ${}_1\tilde{p}_l$ decreases with age l if the original model had that property. **b)** Also show that it increases with θ and coincides with the old one if $\theta = 1$. **c)** Let ${}_1p_l$ be the model of Exercise 6.6.1. Use the program of Exercise 6.6.3 to compute the average, remaining length of life for a twenty-year for $\theta = 1.0, 1.1, 1.2, \dots$ up to $\theta = 2$ and plot the relationship. **d)** Use the plot to find out roughly how large θ must be for the average age to be five years more than it was.

Exercise 6.6.5 Consider a policy holder entering a pension scheme at time $k = 0$ at age l_0 and making a contribution (premium) at the start of each period. From age l_r he draws benefit ζ (also at the start of each period) which lasts until the end of his life. There is a fixed rate of interest r. Let V_k be the value of his account after time k. **a)** Argue that *as long as the member stays alive*, his account develops according to the recursion

$$\begin{aligned} V_k &= (1+r)V_{k-1} + \pi, & k &< l_r - l_0 \\ &= (1+r)V_{k-1} - \zeta, & k &\geq l_r - l_0 \end{aligned} \qquad \text{starting at} \qquad V_0 = \pi.$$

a) Write a program that allows the account to build up and then decline, the scheme terminating upon death. **b)** Simulate and plot the movements of the account against time when

$$l_0 = 30, \qquad l_r = 65, \qquad \pi = ? \qquad \zeta = ? \qquad r = 3\%$$

and the survival model is the one in Exercise 6.6.1. **c)** Repeat b) nine times to judge variability. **d)** If you apply the program ?? on ?? under the Cambrige website you can see how much the status of the account varies when the scheme stops at the death of the policy holder. The plot is based on 10000 simulations under the conditions above.

Section 6.7

Exercise 6.7.1 a) Show that when U_1 is uniform and $U_2 = U_1$, then

$$H^{\text{ma}}(u_1, u_2) = \min(u_1, u_2), \qquad 0 \leq u_1, u_2 \leq 1.$$

is the copula for the pair (U_1, U_2). **b)** Prove the first half of the **Frechet-Hoeffding** inequality; i.e.

$$H(u_1, u_2) \leq \min(u_1, u_2), \qquad 0 \leq u_1, u_2 \leq 1.$$

for an *arbitrary* copula $H(u_1, u_2)$. This shows that $H^{\rm ma}(u_1, u_2)$ is a *maximum* copula.

Exercise 6.7.2 The second half of the Frechet-Hoeffding inequality apply to antitetic variables, introduced in Chapter 4 to produce negatively correlated random variables. Let U_1 be uniform and $U_2 = 1 - U_1$. **a)** Show that the copula is

$$H^{\rm mi}(u_1, u_2) = \max(u_1 + u_2 - 1, 0),$$

For an *arbitrary* copula $H(u_1, u_2)$ fix u_2 and define the function

$$G(u_1) = H(u_1, u_2) - (u_1 + u_2 - 1).$$

b) Show that $G(1) = 0$ and that $G'(u_1) < 0$ [Hint: Recall (**??**).]. **c)** Explain that this means that $G(u_1) > 0$ so that

$$H(u_1, u_2) \geq \max(u_1 + u_2 - 1, 0),$$

and the antitetic pair defines a *minimum* copula.

Exercise 6.7.3 We might use the the preceding two exercises used to check whether a family of copulas capture the entire range of dependency. **a)** Show that the Clayton copula (1.42) coincides with the minimum (antitetic) copula when $\theta = -1$ and **b)** that it converges to the maximum copula as $\theta \to \infty$ [Hint: Utilize that the Clayton copula for $\theta > 0$ may be written

$$\exp\{L(\theta)\} \qquad \text{where} \qquad L(\theta) = \log(u_1^{-\theta} + u_2^{-\theta} - 1)/\theta$$

and apply l'Hôpital's rule to $L(\theta)$.].

Exercise 6.7.4 Show that the Clayton copula (1.40) approaches the independent copula as $\theta \to 0$ [Hint: Use the argument of the preceding exercise.].

Exercise 6.7.5 One of the most popular copula models is the **Gumbel** family for which

$$H(u_1, u_2) = \exp\{-Q(u_1, u_2)\} \quad \text{where} \quad Q(u_1, u_2) = \{(-\log u_1)^\theta + (-\log u_2)^\theta\}^{1/\theta}.$$

a) Verify that this is a valid copula when $\theta \geq 1$ by checking (1.39). **b)** Which model corresponds to the special case $\theta = 1$? **c)** Which model appears as $\theta \to \infty$? [Hint: One way is to utilize that

$$Q(u_1, u_2) = \exp\{L(\theta)\} \qquad \text{where} \qquad L(\theta) = \log[(-\log u_1)^\theta + (-\log u_2)^\theta]/\theta\}.$$

Apply l'Hôpital's rule to $L(\theta)$.]

Exercise 6.7.6 Show that the Gumbel family of the preceding exercise belongs to the Archimedean class with generator $\phi(u) = (-\log u)^\theta$.

Exercise 6.7.7 Let $H(u_1, u_2) = \phi^{-1}\{\phi((u_1) + \phi(u_2)\}$ be a general Archimedean copula where it is assumed that the generator $\phi(u)$ decreases continuously from infinity at $u = 0$ to zero at $u = 1$. **a)** Calculate $H(u_1, 0)$ and $H(0, u_2)$ and verify that the first line in (1.39) is satisfied. **b)** Same question for the second line and $H(u_1, 1)$ and $H(1, u_2)$.

Exercise 6.7.8 Consider the Archimedean copula based on the generator $\phi(u) = (1 - u)^3$. Derive an expression $H(u_1, u_2)$ and b) show that it is zero whenever $u_2 \leq \{1 - (1 - u_1)^3\}^{1/3}$.

Exercise 6.7.9 Suppose an Archimedean copula is based on a generator for which $\phi(0)$ is *finite*. Use the fact that the generator is strictly decreasing to explain that the copula $H(u_1, u_2)$ is positive if and only if

$$\phi(u_1) + \phi(u_2) < \phi(0) \quad \text{true if and only if} \quad u_2 > \phi^{-1}\{\phi(0) - \phi(u_1)\},$$

and the lower bound on u_2 is normally positive. We rarely want models with this property.

Exercise 6.7.10 Consider the Clayton copula (**??**) with positive θ with generator $\phi(u) = (u^{-\theta} - 1)/\theta$. Show that the key part of Algorithm 6.4 (lines 2 and 3) is solved by

$$U_2^* = \{1 + (U_1^*)^{-\theta}[(V^*)^{-\theta/(1+\theta)} - 1]\}^{-1/\theta}.$$