1 Modelling II: Conditional and non-linear

1.1 Introduction

Insurance risk requires modelling tools different from those of the preceding chapter. Pension insurance makes use of **life cycle** descriptions of individuals. They 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 possibly be constructed by means of linear relationships like in the preceding chapter. There are no numerical variables to connect! Instead we link distributions.

The central concept is **conditional** probabilities, 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 that 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

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. It is assumed that you are familiar with conditional probabilities at an introductory level (if not, there is a brief section in Appendix A). When an event A has occurred, the probability of another one B changes from $\Pr(B)$ to

$$\Pr(B|A) = \frac{\Pr(A \cap B)}{\Pr(A)}, \tag{1.1}$$

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. For conditional density functions and conditional expectatons this reads $f(y|x)$ and $E(Y|x)$ if $X = x$ is given.

Modelling can only be learnt by example, and the present section is a bunch of cases. We start with bivariate normal models. These are important in themselves, and introduce the main concepts nicely:

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

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

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

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

called the **survival probability**. This is the likelihood that a person of age y_0 lives at least t longer. If $F(y)$ is the distribution function of Y, then from (1.1)

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

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 $t = kh$. The probability of surviving the coming k time steps must be equal to

$$\underset{}{{}_kp_{l_0}} \;=\; \underset{\textit{first interval}}{{}_1p_{l_0}} \;\times\; \underset{\textit{second interval}}{{}_1p_{l_0+1}} \;\times\cdots\times\; \underset{\textit{k'th interval}}{{}_1p_{l_0+k-1}}, \tag{1.6}$$

and survival modelling is built up from the one-step probabilities ${}_1p_l$; see Section 3.4 for a specific example.

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

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

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

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

where $F(y)$ is the distribution function of Y. 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 $f(y)$ is *not* identically zero above some upper limit, then $f_b(z)$ become either a *Pareto* density or an *exponential* one[1] as $b \to \infty$. This applies no matter which distribution we started with and suggests Pareto models for extreme tails ; see Chapter 9.

Risk hetereogenity
It was in Chapter 3 suggested that random variation for claim frequency N in property insurance should be described by $(n = 0, 1, \ldots)$

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

see (**??**)and (**??**). The central parameter is μ, the claim intensity. Why should that quantity necessarily be the same for everybody? In automobile insurance where drivers are of different ability, there must be discrepancies. Neither are general conditions necessarily the same in all periods. Weather influencing driving is an example, in some countries causing considerable variation; see Chapter 8.

Modelling is the same whether μ affects policies *individually* or the entire portfolio *collectively*.

[1]The exponential model is a limiting member of the Pareto class; see Section 2.6 and the limit is therefore a member of a generalised Pareto class.

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 $\mathcal{N}$ through a conditional model given μ; i.e.

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

Clearly Z is positive, and we should impose $E(Z) = 1$ to make ξ the *mean intensity*. The standard model for Z is Gamma(α), one of the distributions introduced in Section 2.6. Then

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

and the variability in μ, controlled by α, is removed when $\alpha \to \infty$. In the limit μ becomes fixed as ξ.

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 same feature might also apply to *sizes* of claims, and in CAPM models (Section 5.3) the market component played that role. 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 frequency over 25 years were generated 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 seem to match this fairly well[2], but the main point is the uncertainty which is relative terms is no smaller for the large portfolio. That runs contrary to what has seen before (Section 3.2) and reflects that the effect of common factors isn't removed through 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

[2]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 on two.

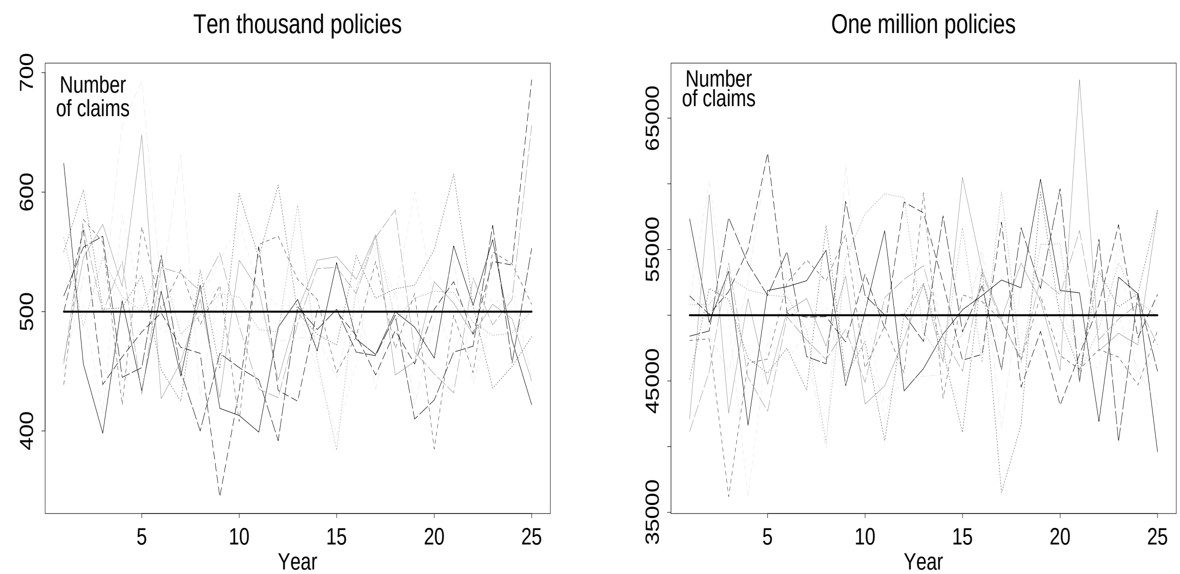


Figure 6.1 *Simulated portfolio claim frequency scenarios under annual change of risk*

ordinary Monte Carlo randomness. To be specific, suppose claim frequency $\mathcal{N}$ against a portfolio follows the ordinary Poisson model and let $\hat{\mu}$ be the estimated claim intensity (estimation method in Chapter 8). The scheme is then

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

and the question is how we examine the impact of both sources of error. A first step is to notice that the model for $\mathcal{N}^*$ really is 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

Situations where a risk variable X is influenced by a second random factor ω on a **subordinate** level were introduced in the preceding section. 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 through a looser specification which makes use of 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 what impact it has on portfolio risk that the mean and standard deviation vary with ω.

The double rules

Our tool is two operational rules that is best introduced in terms of a general random variable Y dependent on a random vector $\mathbf{X}$ of an arbitrary number of variables. 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 $\sigma(\mathbf{x}) = \mathrm{sd}(Y|\mathbf{x})$ influence Y. Important insight is provided by the identities

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

and

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

which are proved in Appendix A. Note that $\xi(\mathbf{X})$ and $\sigma^2(\mathbf{X})$ are random variables. Their expectation and variance lead to the expectation and variance of Y. The double variance formula decomposes $\mathrm{var}(Y)$ into two positive contributions and has consequences reaching far.

These formulae do not require 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 integrated.

Impact of subordinate risk

Let $X_1, \dots, X_J$ be risk variables with X_j depending on a subordinate factor ω_j. These are still seen as independent risks, but this now means that they are **conditionally independent** given $\omega_1, \dots, \omega_J$. We shall consider the two different sampling regimes

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

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 portfolio risk $\mathcal{X} = X_1 + \ldots + X_J$ is widely different, as we shall now see. It will be assumed that all risks X_j and all subordinate factors 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 $\sigma(\omega) = \mathrm{sd}(X_j|\omega)$ are the same for all j. Hence, by adding all contributions

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

the variance formula demanding 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

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

and 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 j, we obtain mean and variance on portfolio level; i.e

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

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 $\mu_1 \dots, \mu_J$ influence the claim frequency $\mathcal{N} = N_1 + \dots + N_J$ of the portfolio under the two sampling regimes above:

$$\underset{common\ factor}{\mu_1 = \dots = \mu_J = \mu} \qquad \text{and} \qquad \underset{individual\ parameters}{\mu_1, \dots, \mu_J \quad \text{independent.}}$$

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 $N_1, \dots, N_J$ are in either case conditionally independent and Poisson given $\mu_1, \dots, \mu_J$. In particular, for an arbitrary N

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

which are the functions $\xi(\mu)$ and $\sigma^2(\mu)$ in (1.15) and (1.16). With $\xi_\mu = E(\mu)$ and $\sigma_\mu = \mathrm{sd}(\mu)$ as the mean and standard deviation of μ (1.15) yields

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

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

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

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

and the standard deviation has the familiar form proportional to $\sqrt{J}$. The practical significance of such risk hetereogeneity over policies 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

$$\mathcal{X} = \sum_{i=1}^{\mathcal{N}} Z_i$$

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

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

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

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

so that

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

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

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

which will be used repeatedly.

Random claim intensity: Important at portfolios level?
We may examine how portfolio liabilities $\mathcal{X}$ are affected by random claim intensities by inserting the expressions in (1.17) and (1.18) for $E(\mathcal{N})$ and $\mathrm{sd}(\mathcal{N})$ into (1.19). For the mean this yields

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

the same whether μ is generated as a common value for the entire portfolio or individually for each policy. That is different with the standard deviation, but a little algebra (detailed in Section 6.8) leads to

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

where

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

This lengty expression tells a lot. The factor $\sqrt{1 + \delta\gamma}$ on the very right in (1.22) is caused by the uncertainty in μ and makes portfolio go up.

But by how much? In practice δ is quite small (hardly more than a few per cent, see Exercise 6.3.2). This leads to the following observation. Suppose $\mu_1, \ldots, \mu_J$ are drawn independently of each other. Then

$$\sqrt{1+\delta\gamma} = \sqrt{1+\delta} \doteq 1 + \delta/2,$$

not a high increase in risk. Indeed, on portfolio level hetereogenity between policies usually contributes little extra risk, but this changes when μ is a collective risk factor. Now $\sqrt{1+\delta\gamma} = \sqrt{1+J\delta}$ which for large J could 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. If $\mathbf{X}$ is a quantity observed, we might use

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

as a way of guessing the value of an unknown Y. Indeed, in theory this is the best way to do it (see below). The result is a celebrated one 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 conditional means. Suppose $\mathbf{X}$ is information possessed at a certain point in time. Then $E(Y|\mathbf{X})$ is what is expected given that knowledge and could be a natural break-even price for carrying the risk Y. Shouldn't what we charge reflect what we know? For example $\mathbf{X}$ might in property insurance summarize our experience with a policy holder or supply other information with bearing on risk. Both versions will appear in Part II; see Chapter 10 in particular. Here the main example is the pricing of money market products such as bonds. This leads us to the theoretical interest rate curve and the **term structure** for bonds over different times to maturity. Specific solutions are obtained by invoking models from Section 5.6.

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

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. The prediction $\hat{Y}$ is thus **unbiased**; more on that in Chapter 7. On the right $\tilde{Y} = \tilde{Y}(\mathbf{X})$ is an arbitrary function of $\mathbf{X}$, and the inequality shows that the conditional mean is on average the most accurate way of utilizing the information $\mathbf{X}$. The proof is given in Appendix A.

As an example consider interest rates following the Vasicĕk model of Section 5.6. The rate r_k at time t_k can then be written

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

see (**??**). Here r_0 is known at $t_0 = 0$ (current time), and $\varepsilon_1, \varepsilon_2, \ldots$ are independent random disturbances with zero mean and unit variance. It was shown in section 5.6 that

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

On the left $\hat{r}_k = E(r_k|r_0)$ is the best prediction of r_k if the Vasicĕk model is true.

What could the accuracy be? 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 standard deviation becomes 1.4% after one year and 2.2% after five. This signifies huge errors, up to $3 - 4\%$ and more. Forecasting interest levels through simple statistical procedures 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 spot rate of interest today (known) and $r_1, \ldots, r_k$ those of the future (unknown). Introduce

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

Here D_k is a discount and had future rates of interest been known, that is what you would pay for a zero-coupon bond expiring at t_k; i.e for the right to receive one money unit at that time. In practice D_k is unknown, but there are anticipations of what it is going to be. Suppose a description in terms of a stochastic model is available. We may then calculate the conditional mean in (1.26), and a rational financial market turns out to value bonds by it! The result requires mathematics of some depth and is proved in Chapter 14. It will then emerge that the Q-model describing future operations in the money market is not the same as the one underlying the spot rate r_k. We encountered the same dualism with equity in Section 3.5, and there are many common features in the underlying theories.

For the zero-coupon bonds of Section 1.4 there are now two sets of prices The preceding **term structure** $P(r_0, t_k)$ is a theoretical one based on a mathematical description of market view and expectations, and there are also the real prices $P(0{:}k)$ traded. 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 spot rate . 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.

Shouldn't the observed set of bond prices $P(0{:}k)$ and the theoretical ones $P(r_0, t_k)$ at time $t_0 = 0$

equal? They must if the Q-model correctly reflects the market view on future uncertainty, and it is common to calibrate its parameters by matching the two sets.

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

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

written as in (**??**) with h present in the notation. Note that the parameters are subscripted with q to emphasize risk-neutrality. Calculations of (1.26) under this model are carried out in Exercises 5.7.12-16. This leads 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.

Monte Carlo term structures
With modern computational power simple bond price formulae may not be so important as before. Indeed, 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 $P(r_k^*, t_k)$ may then be read off approximately from the table. There *is* a minor numerical inaccuracy, but of little practical importance, and we may now employ any Q-model we like. The following implementation is for the Black-Karisinsky model (for which simple solutions do not exist);

Algorithm 6.1 The Black-Karisinsky term structure

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

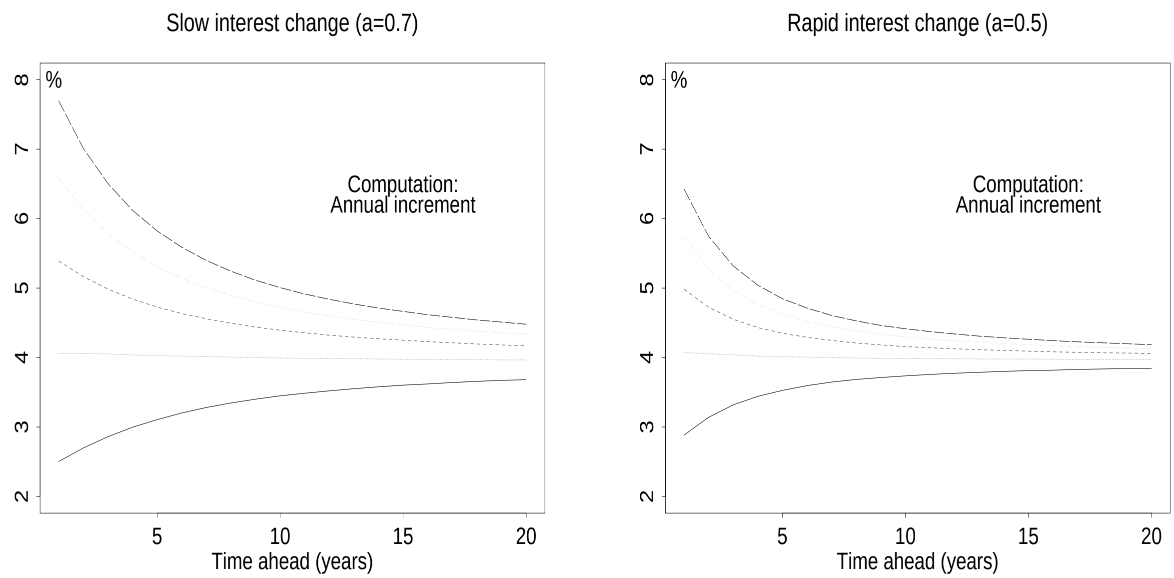


Figure 6.2 *Interest rate curves (from $m = 10000$ simulations) under the Black-Karisinsky model when the initial rate of interest is varied.*

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 5.6. The examples in Figure 6.2 have been run om a crude annual one with parameters

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

and with bond prices 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% and lead to the different shapes in Figure 6.2. In the long run the yield tends to $\xi = 4\%$ as an average 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

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

and $f(x_1, \ldots, x_n)$ is its n-fold partial derivative with respect to $x_1, \ldots, x_n$. In practice we may think of it as the likelihood of the event

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

though formally (in a strict mathematical sense) this probability is zero. Textbooks in probability and statistics often *start* with density functions. They may play a vital role in checking logical consistency in stochastic modelling, but in this book that is always obvious, and we need not go into it. Joint density functions are also needed for the **likelihood** criterion in the next chapter, which often opens for the best possible use of historical data. Copulas in Section 6.7 are examples of modelling joint densities directly.

Factorization of density functions

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

$$\begin{array}{lcccc} \text{Sample} & X_1^* & X_2^*|X_1^* & \cdots & X_n^*|X_1^*, \dots, X_{n-1}^* \\ \text{Probabiltities} & f(x_1) & f(x_2|X_1^*) & \cdots & f(x_n|X_1^*, \dots, X_{n-1}^*), \end{array}$$

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.

Multiplying probabiltities of single events leads to probabiltities of joint events; see Appendix A. Here this exercise leads to the general factorization

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

which reflects that the sampling scheme above produces a Monte Carlo simulation from $f(x_1, \dots, x_n)$. In (1.29) the joint density is broken down on a sequence of conditional ones. Several special cases are of interest.

Types of dependence

The model with a common random factor in Section 6.2 is of the form

$$\underbrace{f(x_1, \dots, x_n) = f(x_1)f(x_2|x_1)\cdots f(x_n|x_1)}_{\textit{Common factor: First variable}}. \tag{1.30}$$

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

$$\underbrace{f(x_1, \dots, x_n) = f(x_1)f(x_2)\cdots f(x_n)}_{\textit{Independence}}. \tag{1.31}$$

Finally, there is the issue of **Markov dependence**, typically associated with time series. Now X_k is recorded at time t_k. The model is

$$\underbrace{f(x_1, \dots, x_n) = f(x_1)f(x_2|x_1)\cdots f(x_n|x_{n-1})}_{\textit{Markov dependence}}, \tag{1.32}$$

where X_k only depends on the preceding X_{k-1}, those before t_{k-1} 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 are summarized in 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 $X_1^*, \dots, X_n^*$

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 Vasicěk model in the form (**??**) which reads ($k = 1, 2, \dots$)

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

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

$$r_k|r_{k-1} = r \quad \sim \quad N(\xi + a(r - \xi), \sigma),$$

and we could iterate over $k = 1, 2 \dots$ as above. No particular benefit over the approach in Chapter 5 results from this. Indeed, the dynamic properties of the model were in Section 5.6 derived without introducing the normal.

The sequence $r_1, \dots, r_k$ of interest rates now become jointly Gaussian (see Section 3.4), and it *is* possible to write down its density function. However, except for model calibration (Chapter 7) we rarely have much use for it. The *general* Gaussian density function reads

$$f(\mathbf{x}) = (|2\pi\Sigma|)^{-1/2} \exp\{-\frac{1}{2}(\mathbf{x} - \boldsymbol{\xi})'\Sigma^{-1}(\mathbf{x} - \boldsymbol{\xi})\} \tag{1.33}$$

where $\boldsymbol{\xi} = (\xi_1, \dots, \xi_n)'$ is the vector of expectations, Σ the covariance matrix and $|2\pi\Sigma|$ the determinant of the matrix. This expression, though famous, plays no role in this book.

Example: The multinomial situation
One joint density function that *will* be used is the multinomial one which will be needed for modelling delays in property insurance in Section 8.5. Multinomial sampling (Section 4.2) means randomly selecting one label among K different ones. With n independent repetitions according to the same probabilities $p_1, \dots, p_K$ we are dealing with a **multinomial** sequence of trials, and the number of times $N_1, \dots, N_K$ the various labels appear has a multinomial distribution. Here $p_1 + \dots + p_K = 1$. With $K = 2$ we are back to the familar binomial situations. Actually the general case can be derived from this special one.

The argument is carried out for $K = 3$. Write bin(n, p) for the binomial distribution based on n trials and success probabiltiy p. Then with $K = 3$

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

the last one (N_K) always being fixed by the $K-1$ others. The conditional distribution of N_2 is as stated because the remaining $n - n_1$ trials are either label 2 or 3. But the joint density function of (N_1, N_2) then 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 factor cancels. This leaves as the joint density

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

The general case is

$$\Pr(N_1{=}n_1, \ldots, N_K{=}n_K) = \frac{n!}{n_1! \ldots n_K!}\, p_1^{n_1} \ldots p_K^{n_K} \qquad (1.34)$$

where $n_1 + \ldots + n_K = n$.

1.6 Markov chains and life insurance

Introduction

Liability risk in life and pension insurance are based on probabilistic descriptions of life cycles, as 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 of thousands (or 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. Do we really need it? After all, uncertainty due to life cycle movements is rarely very important (see Section 3.4). But that doesn't mean that the underlying stochastic model is irrelevant. It is needed both to compute the expectations defining the liabiltities *and* to evaluate portfolio uncertainty due to errors in parameters.

Markov modelling

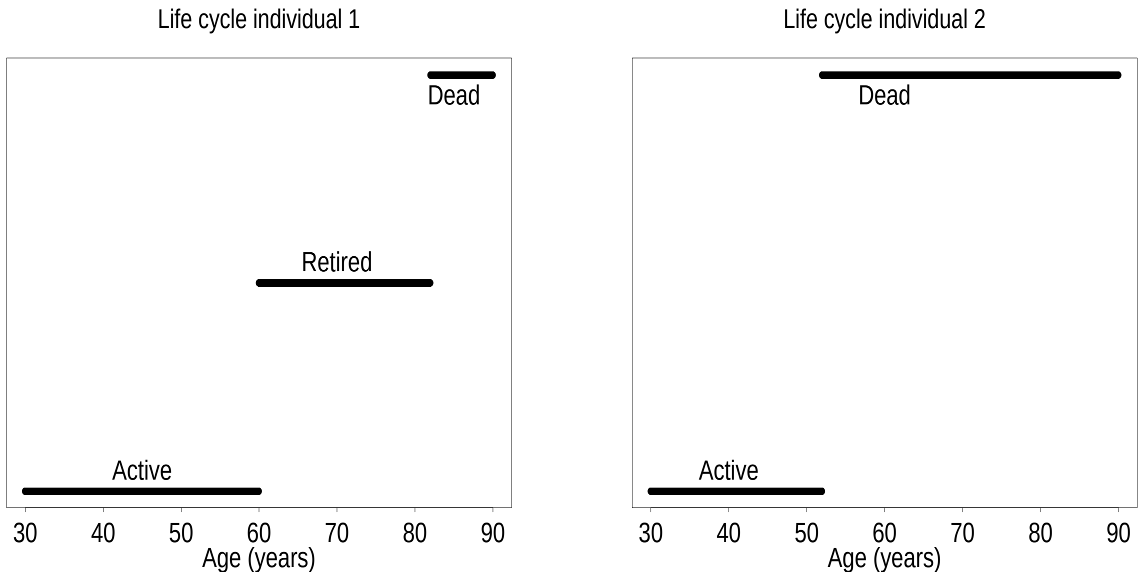


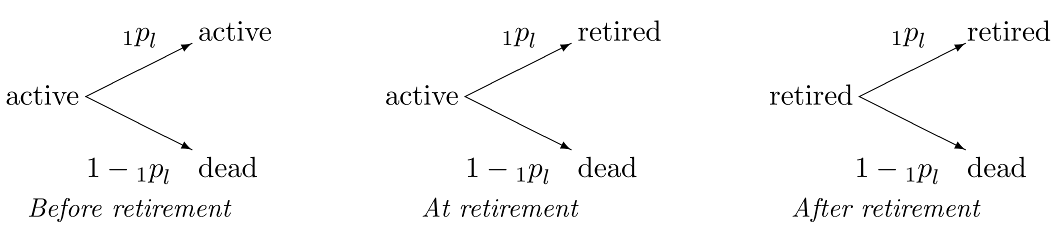
Figure 6.3 The life cycles of two members of a pension scheme.

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**

$$p_l(i|j) = \Pr(C_{l+1} = i | C_l = j). \tag{1.35}$$

Algorithm 6.2 tells us how life cycles governed by a Markov chains develop. At each point in time there is a random experiment taking 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 understand how such models work; see Exercises 6.6.2 and 6.6.5.

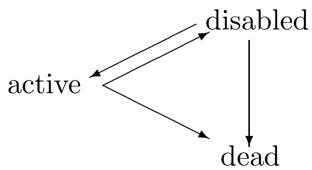
Transition probabilities are usually different for men and women (not reflected in the mathematical notation), and it is (of course) essential that they depend on age l. A major part of them always come from the survival probabilities ${}_1p_l$ introduced in Section 6.2. For a simple pension scheme,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



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 (2005), 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 $p_{a|i}$. The former is the probability of moving from 'active' to 'disabled' and the other the opposite.

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

	To new state			
From	Active	Disabled	Dead	Row sum
Active	${}_1p_l \cdot (1 - p_{i\|a})$	${}_1p_l \cdot p_{i\|a}$	$1 - {}_1p_l$	1
Disabled	${}_1p_l \cdot p_{a\|i}$	${}_1p_l \cdot (1 - p_{a\|i})$	$1 - {}_1p_l$	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 discover that it is true). *Any* set of transition probabilities for Markov chains must satisfy this restriction, which merely reflects that the individual always moves somewhere or remains in the same state.

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.

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

Their corresponding annual **mortalitites** ${}_1q_l = 1 - {}_1p_l$ 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 2% 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

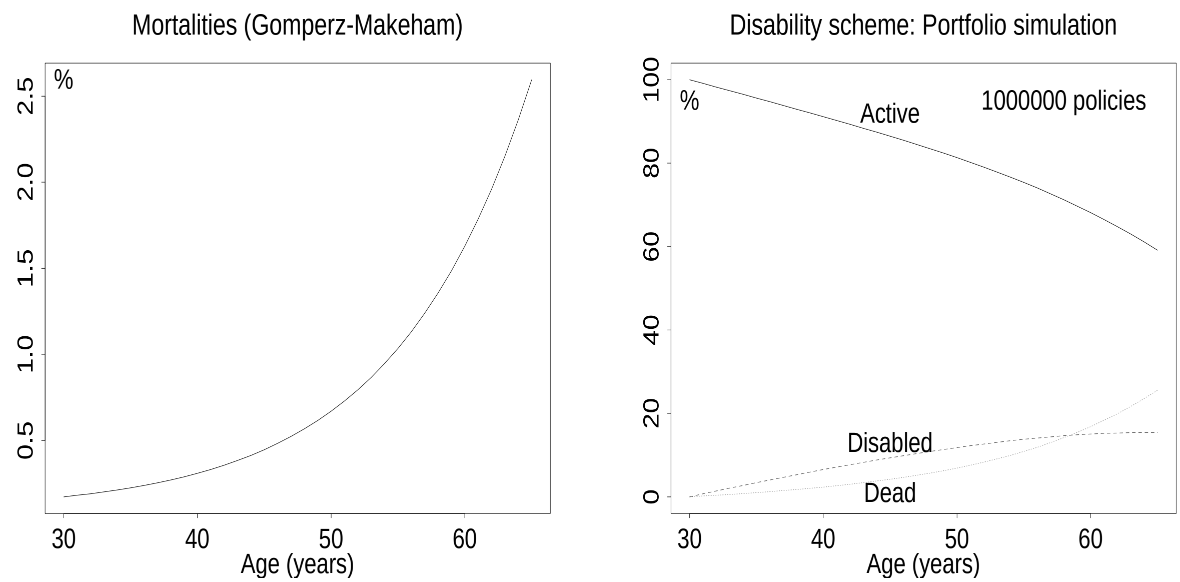


Figure 6.4 A disability sceme in life insurance: Mortality model (left) and portfolio simulation (right).

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

$$p_{i|a} = 0.7\%, \qquad \text{and} \qquad p_{a|i} = 0.35\%,$$

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

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 some 75% remain alive, a realistic figure. What is not necessarily true in practice is the downwards curvature in the disability curve which might have been upwards if the disability rate is made age-dependent.

1.7 Introducing copulas

Introduction

The copula concept is an old one, going back to the mid twentieth century. Yet it is only in fairly recent years it has attracted interest as a tool for actuarial and financial risk. An early contribution is Carriere (1987). The idea has much to do with sampling by inversion; see Chapter 2. Let X_1 and X_2 be random variables with strictly increasing distribution functions $F_1(x_1)$ and $F_2(x_2)$. Then

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

where U_1 and U_2 are uniformly distributed. They do not have to be independent which is precisely where coupla modelling comes in Dependence are now formulated in terms of U_1 and U_2, which are then mapped back to X_1 and X_2. Note that the dependence now has become a modelling issue *completely detached* from the distributions of X_1 and X_2. The power of this idea will emerge below. *All* bivariate and (more generally multivariate) stochastic models can be represented in this way.

Copulas differ from the other approaches to modelling in this chapter in that it is *non-constructive*. The way it is defined does *not* give a simple recepy for how such models are simulated in the computer. That is an area begging for development. What is available has influenced the way this section has been written. One model with attractive theoretical properties and at the same time easy to simulate is the **Clayton** family. This is one of the most frequently applied copulas, member of the **Archimedean** class, also widely used. Most of the section is devoted to those. We start bivariately and extend to J variables at the end.

What is a copula?

A **copula** is a joint distribution function for *dependent* uniform random variables. In the bivariate case this means the function

$$C(u_1, u_2) = \Pr(U_1 \le u_1, U_2 \le u_2), \tag{1.36}$$

defined for all u_1 and u_2 between 0 and 1. For a valid model we must require $C(u_1, u_2)$ to be *non-decreasing* in both u_1 and u_2 and

$$\begin{aligned} &C(u_1, 0) = 0, \qquad C(0, u_2) = 0 \\ &C(u_1, 1) = u_1, \quad\ C(1, u_2) = u_2 \end{aligned} \tag{1.37}$$

for all u_1 and u_2. For example

$$C(u_1, 1) = \Pr(U_1 \le u_1, U_2 \le 1) = \Pr(U_1 \le u_1) = u_1,$$

and similar for the others. Any function $C(u_1, u_2)$ serving as a copula must satisfy (1.37).

The most immediate example is

$$C(u_1, u_2) = u_1 u_2$$

making U_1 and U_2 independent. More interesting is the **Clayton** copula for which

$$C(u_1, u_2) = (u_1^{-\theta} + u_2^{-\theta} - 1)^{-1/\theta}, \quad 0 < u_1, u_2 < 1, \qquad (\theta > 0) \tag{1.38}$$

Here θ is a positive parameter (negative ones will be allowed later). It is easily verified that (1.37) is satisfied. The independent copula appears in the limit as $\theta \to 0$; see Exercise ?. The Clayton model has a number of attractive properties and is one of the most useful couplas.

Copula modelling

The previous discussion has suggested the following modelling strategy. Start by finding appropriate distribution functions $F_1(x_1)$ and $F_2(x_2)$ for X_1 and X_2 and then throw a copula $C(u_1, u_2)$

around them to account for dependency. From what was said above the joint distribution function for the pair (X_1, X_2) becomes

$$\underbrace{F(x_1,x_2) = C(u_1,u_2)}_{\textit{copula modelling}} \qquad \text{where} \qquad \underbrace{u_1 = F_1(x_1), \;\; u_2 = F_2(x_2).}_{\textit{univariate modelling}} \tag{1.39}$$

This is actually a general representation, discovered by **Sklar** (1959) and bears his name. *Any* bivariate distribution function $F(x_1, x_2)$ can be written in this form, provided the marginal distribution functions $F_1(x_1)$ and $F_2(x_2)$ are strictly increasing. A modified version holds for counts[3], and Sklar's result can be extended to any number of variables.

In (1.39) either of the relationships on the right may be replaced by their anti-tetic twin (see Section 4.5). This produces the three additional versions

$$\begin{array}{ll} u_1 = F_1(x_1), \;\; 1-u_2 = F_2(x_2) & \textit{orientation (1,2)} \\ 1-u_1 = F_1(x_1), \;\; u_2 = F_2(x_2) & \textit{orientation (2,1)} \\ 1-u_1 = F_1(x_1), \;\; 1-u_2 = F_2(x_2) & \textit{orientation (2,2),} \end{array} \tag{1.40}$$

all combined with the *same* copula on the left in (1.39). The effect (see Figure 6.5) is to rotate the copula patterns 90°, 180° and 270° compared to the orginal one which will be called orientation $(1, 1)$.

The Clayton copula

The copula bearing the name of the British statistician David Clayton was introduced above. Its defintion through (1.38) can be extended to include *negative* θ down to -1, provided the mathematical expression is modified to

$$C(u_1, u_2) = \max\{(u_1^{-\theta} + u_2^{-\theta} - 1)^{-1/\theta}, 0\} \qquad (\theta \geq -1). \tag{1.41}$$

Again it is easy to check that the copula requirements (1.37) are satisfied when $\theta \geq -1$. For negative θ, the expression is positive when

$$u_2 > (1 - u_1^{-\theta})^{-1/\theta}.$$

Below that threshold the copula is zero; see also Figure 6.6 right. Usually restrictions of that kind are undesirable. Still, when the negative part is incuded, the family in a sense cover the entire range of dependency that is logically possible; see Exercises ?? and ??.

Examples of structures generated by the Clayton copula are shown in Figure 6.5. The two marginal distributions were normal with mean $\xi = 0.005$ and volatility $\sigma = 0.05$, precisely as in Figure 3.3. Most striking is the cone-shapes patterns which signify unequal degree of dependence in unequal parts of the space. Consider, for example, the plot in the upper, left corner where correlation in downside returns are much stronger than for upside ones. Such phenomena have been detected in practice; see Longin and Solnik (2002). Consequences for downside risk could be serious. Ordinary Gaussian models can't capture this. The other plots in Figure 6.5 rotate patterns by varying the orientation of the copula. Dependence is adjusted by moving the Clayton parameter θ (high values

[3]The distribution functions are then not *strictly* increasing, as demanded by the theorem.

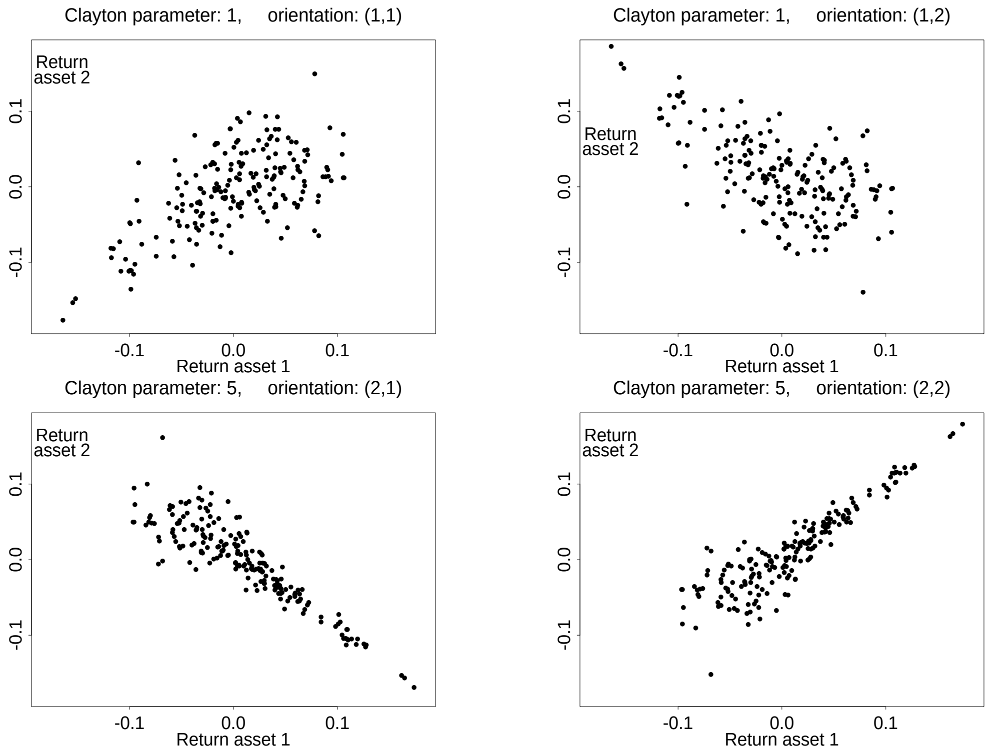


Figure 6.5 *Simulated financial returns from normals and* **Clayton** *copula.*

for strong dependence).

Conditional distributions for copulas
As elsewhere it is useful to examine the conditional models. Let

$$C(u_2|u_1) = \Pr(U_2 \leq u_1|u_1)$$

be the conditional distribution function of U_2 given $U_1 = u_1$. This turns out to be the partial derivative of the orginal copula with respect to u_1, i.e.

$$C(u_2|u_1) = \frac{\partial C(u_1,u_2)}{\partial u_1}; \tag{1.42}$$

see Section 6.8.

For the Clayton copula (1.38) straightforward differentiaton yields

$$C(u_2|u_1) = u_1^{-(1+\theta)}(u_1^{-\theta} + u_2^{-\theta} - 1)^{-(1+1/\theta)}, \tag{1.43}$$

where

$$\begin{array}{ll} 0 < u_2 < 1 & \text{for} \quad \theta > 0, \\ (1 - u_1^{-\theta})^{-1/\theta} < u_2 < 1 & \text{for} \quad -1 \leq \theta < 0 \end{array}$$

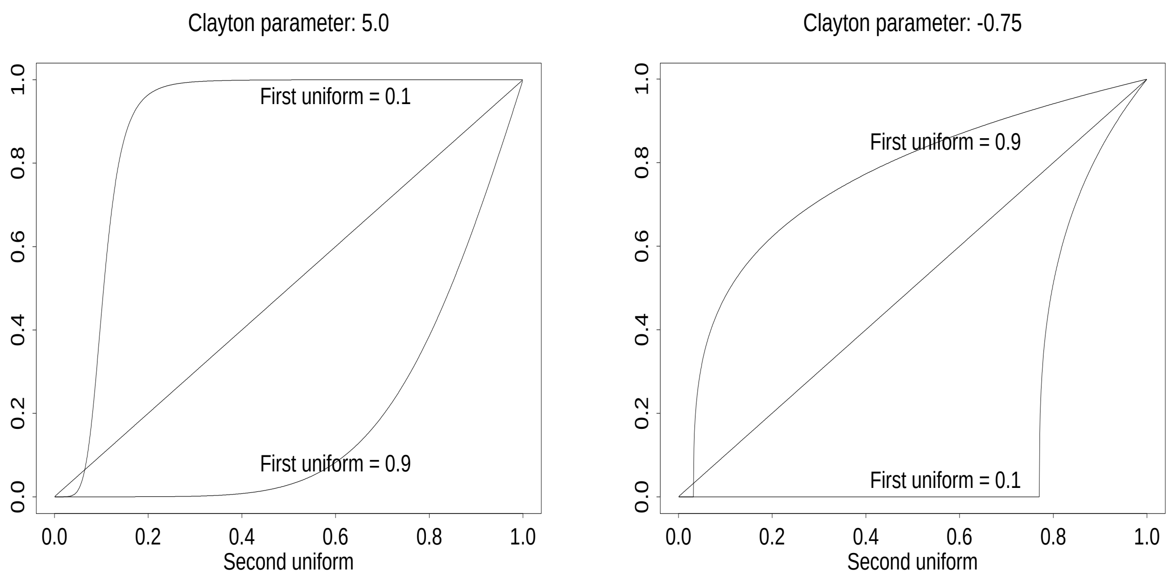


Figure 6.6 Conditional distribution functions for the **second** variable of a Clayton copula; given first variable marked on each curve.

Below the lower threshold $C(u_2|u_1) = 0$. The conditional distribution functions have been plotted in Figure 6.6 for θ large and *positive* on the left and large and *negative* on the right. Shapes for $u_1 = 0.1$ and $u_2 = 0.9$ differ markedly, attesting to strong dependency, but the most notable feature is a lack of symmetry. Consider the distributions on the left. When $u_1 = 0.1$, the second variable U_2 is located in a narrow strip around that value, (i.e. *very* strong correlation), but if $u_1 = 0.9$, the range of variation for U_2 is much larger. It is precisely this feature that creates the cones in Figure 6.5; see also Exercise ?? and ??.

How copula models can be simulated

The most obvious way of sampling copulas is to combine conditional sampling and inversion, as follows:

Algorithm 6.3 Bivariate copulas
0 Input: The conditional copula $C(u_2|u_1)$
1 Draw U_1^* and $V^* \sim$ uniform
2 Determine U_2^* from

$$C(U_2^*|U_1^*) = V^* \qquad \text{\%}\textit{Equation, often demanding a numeric solution}$$

3 Return U_1^* and U_2^*.

The second step is an application of the inversion algorithm, and here there is a problem. For most copulas analytical solutions do not exist, and a numerical procedure has to be used. This obstacle isn't insurmountable, but it does slow the procedure down, especially when there are more than two variables. The Clayton copula is an exception. It is easy to see that the distribution function (1.43) admits an easy solution. The details are worked out at the end of the section.

Output from Algorithm 6.2 must be combined with inversion to generate the original variables X_1 and X_2. Details depend on the orientation. The two most important ones are

$$X_1^* = F^{-1}(U_1^*)\ \ X_2^* = F_2^{-1}(U_2^*) \quad \text{and} \quad X_1^* = F^{-1}(1-U_1^*),\ X_2^* = F_2^{-1}(1-U_2^*).$$

For other possibilites; see Exercise ??.

Archimedean copulas

Perhaps the most important *general* class of copulas is the **Archimedean** one where

$$C(u_1,u_2) = \phi^{-1}\{\phi(u_1)+\phi(u_2)\}. \tag{1.44}$$

Here the function $\phi(u)$ is a so-called **generator**. The Clayton copula is a special case. Its generator and generator inverse are

$$\phi(u) = \frac{1}{\theta}(u^{-\theta}-1), \qquad \text{and} \qquad \phi(x)^{-1} = (1+\theta x)^{-1/\theta}, \tag{1.45}$$

where the inverse is found by solving $\phi(u) = x$ for u. If these expressions are inserted into (1.44), the earlier expression for the Clayton copula emerges.

The Clayton generator is plotted in Figure 6.7 left for $\theta = 0.2$. It is

- *strictly decreasing* and *continous,*
- with $\phi(1) = 0$ and becomes *infinite* as $u \to 0$.

These ensure that the conditions (1.37) are satisfied. The other example in Figure 6.7 is

$$\phi(u) = (1-u)^3, \qquad 0 < u < 1,$$

an example of a **polynomial** copula. It satisfies all the conditions above with one exception. As $u \to 0$ it remaining *finite*. A valid copula is still defined (Exercise 6.7.7), but it inevitably leads to models where certain combinations of u_1 and u_2 are forbidden. Clayton copulas based on *negative* θ) have the same property, and usually we do not want it. It is avoided if the generator is infinite at the origin. Nelson (1997) lists many possibilities.

Both examples in Figure 6.7 are *convex* (curvature upwards). This is a natural additional condition. The derivative $\phi'(u)$ is then increasing and possesses and inverse. It follows (Section 6.8) that the equation in Algorithm 6.3 can be solved producing a simple sampling algorithm:

Algorithm 6.4 Archimedean copulas
0 Input: Convex generator $\phi(u)$.
1 Draw U_1^* and $V^* \sim$ uniform
2 $Y^* \leftarrow \phi'^{-1}\{\phi'(U_1^*)/V^*\}$ % *Note:* $\phi'(u)$ *the derivative of* $\phi(u)$
3 $U_2^* = \phi^{-1}\{\phi(Y^*) - \phi(U_1^*)\}$
4 Return U_1^* and U_2^*

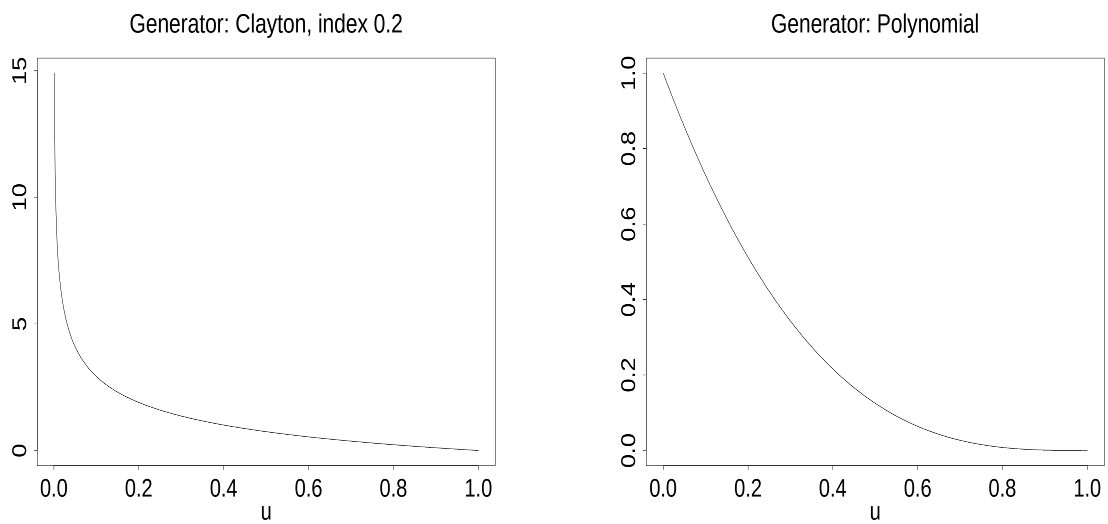


Figure 6.7 Generator functions for Archimedean copulas

If $\phi'^{-1}(x)$ is difficult to find, it can be tabulated on a tight set of points prior to running the algorithm. Table methods for sampling were discussed in Section 4.3.

Copulas with many variables
Some of the ideas and results above extend to J variables without much effort. A J-dimensional copula $C(u_1, \ldots, u_J)$ is the joint distribution function of J dependent uniform random variables $U_1, \ldots, U_J$. Mathematical conditions similar to (1.37), but more complex have to be satisfied. There is an Archimedean type which is an immediate extension of (1.44); i.e.

$$C(u_1, \ldots, u_J) = \phi^{-1}\{\phi(u_1) + \ldots + \phi(u_J)\}. \tag{1.46}$$

Here $\phi(u)$ is a generator of exactly the same type as in the bivariate case. The J uniforms are mapped back to the J original variables $X_1, \ldots, X_J$ through the J inversions

$$X_1 = F_1^{-1}(U_1), \quad \ldots,, \quad X_J = F_J^{-1}(U_J).$$

There are now 2^J ways to rotate patterns through use of anti-tetic twins, not just 4.

Archimedean copulas are still convenient to sample, but a *general* extension of Algorithm 6.4 involves complex chains of derivatives of high order, beyond what is natural to include. For the Clayton copula matters simplify. The following sampling algorithm is justified in Section 6.8:

Algorithm 6.5 The Clayton copula for J variables
0 Input: θ
1 $S^* \leftarrow 0$ and $P^* \leftarrow 1$ %*Initializing auxilliary quantities*
2 Draw $U_1^* \sim$ uniform
3 For $j = 2, \ldots, J$ do
4 $\quad S^* \leftarrow S^* + (U_{j-1}^*)^{-\theta} - 1$ %*Updating from preceding uniform*

$P^* \leftarrow P^*(U^*_{j-1})^{1+\theta}/(1+(j-1)\theta)$ %*New update*
Draw $V^* \sim$ uniform
$U^*_j \leftarrow \{(P^*V^*)^{-\theta/(\theta+j-1)} - S^*\}^{-\theta}$ %*Next uniform*

Return $U^*_1, \ldots, U^*_J$

This algorithm has been used for copula simulations in this book.

1.8 Mathematical arguments

Section 6.3
Portfolio risk We shall verify the formula (1.22) for the standard deviation of the portfolio risk $\mathcal{X}$ starting 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). Expressions for $E(\mathcal{N})$ and $\mathrm{var}(\mathcal{N})$ were given in (1.17) and (1.18). Those are

$$E(\mathcal{N}) = J\xi_\mu T \qquad \text{and} \qquad \mathrm{var}(\mathcal{N}) = JT^2(\gamma\sigma_\mu^2 + \xi_\mu/T)$$

where $\gamma = J$ or $\gamma = 1$ for common and independent sampling of the intensities. Inserting into the expression 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

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

which is (1.22).

Section 6.7.
Conditional distributions for copulas Recall that

$$C(u_1, u_2) = \int_0^{u_1}\int_0^{u_2} h(v_1, v_2)\, dv_1\, dv_2,$$

writing $c(u_1, u_2)$ for the (joint) density function. Hence

$$\frac{\partial C(u_1, u_2)}{\partial u_1} = \int_0^{u_2} c(u_1, v_2)\, dv_2,$$

or, since $c(u_1) \equiv 1$ is the density for U_1

$$\frac{\partial C(u_1, u_2)}{\partial u_1} = \int_0^{u_2} c(v_2|u_1)\, dv_2,$$

where $c(u_2|u_1)$ is the conditional density of U_2. This is (1.42).

Justifying Algorithm 6.4 An Archimedean copula $C(u_1, u_2)$ is according to (1.44) defined through

$$\phi\{C(u_1, u_2)\} = \phi(u_1) + \phi(u_2).$$

When both sides are differntiated with respect to u_1, it follows by the chain rule that

$$\phi'\{C(u_1, u_2)\}\frac{\partial C(u_1, u_2)}{\partial u_1} = \phi'(u_1).$$

so that

$$C(u_2|u_1) = \frac{\phi'(u_1)}{\phi'\{C(u_1, u_2)}$$

is the conditional distribution function of U_2. It follows by the inversion algorithm that a simulation U_2^* is the solution of the equation

$$\frac{\phi'(U_1^*)}{\phi'\{C(U_1^*, U_2^*)} = V^*$$

where U_1^* and V^* are independent and uniform. This may equivalently be written

$$C(U_1^*, U_2^*) = Y^* \qquad \text{where} \qquad Y^* = \phi'^{-1}\{\phi'(U_1^*)/V^*\},$$

which is the quantity on line 2 in Algorithm 6.4. Hence

$$\phi(U_1^*) + \phi(U_2^*) = \phi(Y^*)$$

and when this is solved for U_2^*, Algorithm 6.3 follows.

Justifying Algorithm 6.5 When the Clayton generator and inverse (1.45) are inserted into (1.46), it follows that the mathematical expression for the J-dimensional Clayton copula is

$$C(u_1, \dots, u_J) = \{\sum_{j=1}^{J} u_j^{-\theta} - (J-1)\}^{-1/\theta}.$$

We shall find the conditional distribution function for U_J given the $J-1$ others which equals the partial derivative with respect to $u_1, \dots, u_{J-1}$; i.e.

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

It is straightforwardly derived that

$$\frac{\partial^{J-1} C(u_1, \dots, u_J)}{\partial u_1 \dots \partial u_{J-1}} = \{\sum_{j=1}^{J} u_j^{-\theta} + J - 1)\}^{-(1/\theta+J-1)} \times \prod_{j=1}^{J-1}\{u_j^{-(1+\theta)}(1+(j-1)\theta)\}$$

or

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

where

$$s_{J-1} = \sum_{j=1}^{J-1} u_j^{-\theta} - (J-1) \qquad \text{and} \qquad p_{J-1} = \prod_{j=1}^{J-1} \{u_j^{1+\theta}/(1+(j-1)\theta)\}$$

Suppose $U_1^*, \ldots, U_{J-1}^*$ have been generated and V^* is another uniform, drawn indepdently. Applying inversion we must solve with respect to U_J^* the equation

$$\frac{\partial^{J-1} C(U_1^*, \ldots, U_J^*)}{\partial u_1 \ldots \partial u_{J-1}} = V^*.$$

The Monte Carlo versions of s_{J-1} and p_{J-1} are

$$S_{J-1}^* = \sum_{j=1}^{J-1} (U_j^*)^{-\theta} - (J-1) \qquad \text{and} \qquad P_{J-1}^* = \prod_{j=1}^{J-1} \{(U_j^*)^{1+\theta}/(1+(j-1)\theta)\},$$

and the equation for U_J^* becomes

$$\{(U_J^*)^{-\theta} + S_{J-1}^*\}^{-(1/\theta+J-1)}/P_{J-1}^* = V^*$$

with solution

$$U_J^* = \{(P_{J-1}^* V^*)^{-\theta/\{1+\theta(J-1)\}} - S_{J-1}^*\}^{-1/\theta}.$$

This is how the J'th uniform is generated from the $J-1$ preceding ones. Algorithm 6.5 makes use of this procedure for $J = 2, 3, \ldots$ updating the auxilliary quantities S_{J-1}^* and P_{J-1}^* recursively as we go along.

1.9 Further reading

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

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

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 $\mathrm{sd}(\mathcal{X})$. 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

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

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

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

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 Poisson$(J\hat{\mu})$. **a)** Use the double rules to prove that

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

b) Recall that $E(\mathcal{N}) = \mathrm{var}(\mathcal{N})$ for a Poisson variable $\mathcal{N}$ whereas $E(\mathcal{N}^*) < \mathrm{var}(\mathcal{N}^*)$. 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

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

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

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

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

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

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

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

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

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

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

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

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

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

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**[4] 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

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

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
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 $N_K^* \leftarrow n - S^*$.

[4]ARCH stands for **auto**regressive, **c**onditional, **h**etereochedastic.

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 (1.34) 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 $l = 20, 25, 30, \ldots$ 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

$$ {}_1\tilde{p}_l = \frac{\theta_1 p_l}{\theta_1 p_l + (1 - {}_1p_l)}, $$

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 $\theta = 1.0, 1.1, 1.2, \ldots$ 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

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

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

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

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 $H^{\mathrm{ma}}(u_1, u_2)$ 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

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

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 (1.42).]. **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 (1.41) 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 (1.39) 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 (1.37). **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 (1.37) 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 (1.38) 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

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