# 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 'disable' (benefits again) and all along they may die. Insurance companies and pension schemes must keep track on and plan for such things since their cash flow is influenced. 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 is often not 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 random uncertainties of entirely different origin and define a **hierarchy**. Driver comes first, and conditional modelling is the natural way to connect them. The same viewpoint is crucial when errors of different origin are examined in the next chapter. 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

$$\text{generate } X^* \qquad \text{and then} \qquad Y^* \text{ given } X^*,$$

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 densities 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 $\eta_1$ and $\eta_2$ are independent and normal $(0, 1)$. see (**??**). Suppose $X_1 = x_1$ is fixed. Then $\eta_1 = (x_1 - \xi_1)/\sigma_1$, which when inserted for $\eta_1$ 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 $\eta_2$ is the only random term and, by definition, $X_2$ is normal with mean and standard deviation

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

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 quantitiy 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)

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

Survival probabilities often apply on increments 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 will be written ${}_kp_{l_0}$ when $y_0 = l_0h$ and $t = kh$. The probability of surviving the coming $k$ time steps must be equal to

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

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 $a$ we seek the distribution of

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

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

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

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

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

as the density function for the amount exceeding a given threshold. Tail distributions of this type possess a remarkable property discovered by Pickand (1975). For most distributions used in practice, precisely if $f(y)$ is *not* identically zero above some upper limit, then $f_a(z)$ become either a *Pareto* density or an *exponential* one[1] as $a \to \infty$. 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 $(n = 0, 1, \ldots)$

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

see (**??**)and (**??**). The central parameter is $\mu$, 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 do *general* conditions from one period to another necessarily stay the same. Weather influencing driving is an example, in some countries causing considerable variation from one year to another; see Chapter 8.

Modelling is the same whether $\mu$ affects policies *indvidually* 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 $\mu$ is drawn randomly and *then* $N$ or $\mathcal{N}$ through a conditional model given $\mu$; 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 $\xi$ the *mean intensity*. The standard model for $Z$ is Gamma($\alpha$), 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 $\mu$, controlled by $\alpha$, is removed when $\alpha \to \infty$. In the limit $\mu$ becomes fixed as $\xi$.

**Common risk factors**

The preceding example is a special case of a more general viewpoint. A random variable $\omega$ is called a **common factor** for $X_1, \ldots, X_J$ if

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

That is precisely the situation when the same random intensity $\mu$ affects all claim frequencies $N_1, \ldots, N_J$. Of course, one could also envisage a common random background influencing *sizes* of claims, and we have already (in Chapter 5) met the idea as the market component in CAPM models. The latter is directly observable whereas the others are not. For example, a random intensity $\mu$ is only felt though its indirect effect on claim frequencies. That is an important distinction. Common factors we do not observe or measure directly are called **hidden** or **latent**.

Whether hidden or not common factors 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 risk, expressed though $\mu$, changed every year and was the same for all policies. Suppose $\mu$ 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.

[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.

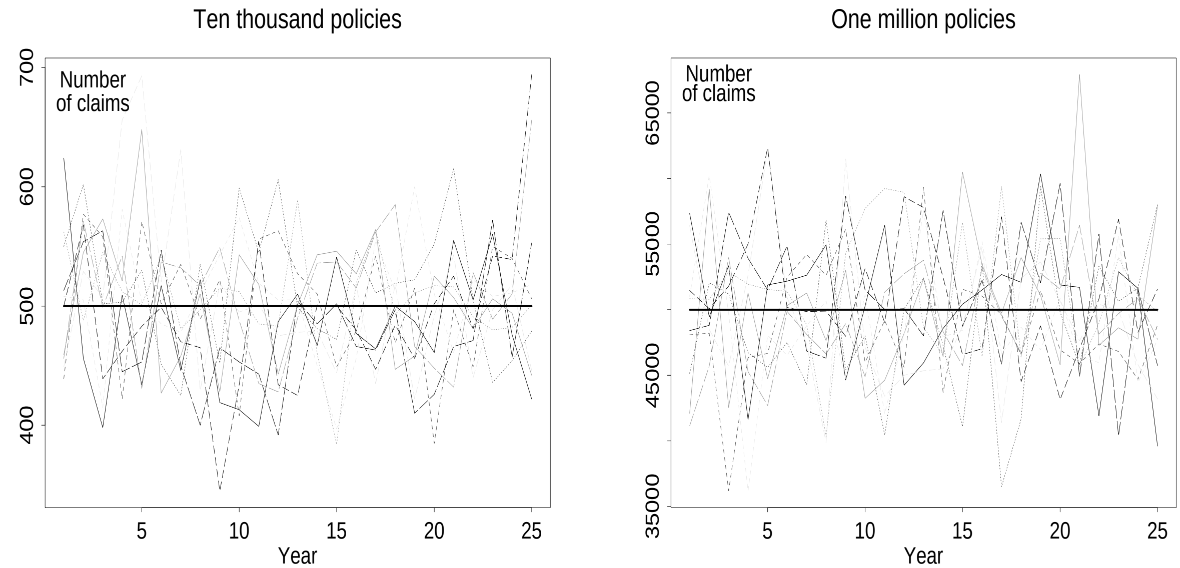



*Figure 6.1 Simulated portfolio claim frequency scenarios under annual change of risk*

**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 claim frequency $\mathcal{N}$ against a portfolio follows the ordinary Poisson model and let $\hat{\mu}$ be the estimated claim intensity (see Chapter 8 for the estimation method). The scheme is then

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

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

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

and we must combine with statistical errors in the estimation process. This is carried out in Chapter 7.

## 1.3 Hierarchic arguments and hidden risk

**Introduction**

Much stochastic modelling is concerned with mean and standard deviation and *conditional* modelling is no exception. Let

$$\xi(x) = E(Y|x) \qquad \text{and} \qquad \sigma(x) = \text{sd}(Y|x) \tag{1.12}$$

be the conditional mean and standard deviation of $Y$ given $X = x$. In the normal case $\xi(x)$ is a straight line and $\sigma(x)$ a constant; see (1.3). Conditional expectation is known as the **regression** of $Y$ on $X$ and leads to **regression analysis** when both $X$ and $Y$ is observed. The familar *linear* one, associated with the normal distribution, is probably not among the most important methods for actuarial science (and is not covered in this book)[3]. Other forms of regression analysis will be introduced in Part II.

This section is principally concerned with situations where $X$ is *not* observed. The issue addressed is how conclusions on $Y$ are drawn from model specifications like (1.12). We are dealing with hierarchic structures where $Y$ is randomly influenced by randomness in $X$. The hidden factors behind risk hetereogenity in the preceding section are examples, and a lot will be said on those. A complete stochastic model is not assumed. Yet it is possible to say quite a lot about the impact of randomness in $X$. Property insurance may be the most fruitful area for the approach, and most of the examples come from there. The mathematical tool is two important operational rules.

**The double rules**
The general formulation replaces $X$ with a random vector $\mathbf{X}$ of an arbitrary number of variables. Then, as is proved in Appendix A,

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

and

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

Note that $\xi(\mathbf{X})$ and $\sigma^2(\mathbf{X})$ both are random variables. We may calculate their expectation and variance, and when they are combined as shown, we end up with the expectation and variance of $Y$. The double variance formula is a decomposition into two (positive) contributions to var($Y$) and has consequences reaching far. A similar rule for double *covariances* is discussed in Exercise 6.3.7.

**Portfolio risk in property insurance**
Aggregated claims in property insurance are easily studied through the double rules. Consider the model from Chapter 3; 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 $\text{sd}(Z_i) = \sigma_z$. 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

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

[3] CAPM and some other models in finance are exceptions and linear regression is one of the key methods in the related field of econometrics.

so that

$$E(\mathcal{X}) = E(\mathcal{N})\xi_z \qquad \text{and} \qquad \text{var}(\mathcal{X}) = E(\mathcal{N})\sigma_z^2 + \text{var}(\mathcal{N})\xi_z^2 \tag{1.15}$$

which are useful formulas. They are applied in Exercise 6.3.1 when $\mathcal{N}$ follows a pure Poisson distribution and put to another use below.

**The effect of common risk**

Common factors were introduced in the preceding section as a model for which $X_1, \ldots, X_J$ are conditionally independent given a random variable $\omega$. Let

$$\xi(\omega) = E(X_j|\omega) \qquad \text{and} \qquad \sigma(\omega) = \text{sd}(X_j|\omega)$$

be the conditional mean and standard deviation and suppose (for simplicity) that they are the same for all $j$. We seek the unconditional mean and standard deviation for their sum $\mathcal{X} = X_1 + \ldots + X_J$. First note that

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

and we may invoke the double rules with $Y = \mathcal{X}$ and $\mathbf{X} = \omega$. By (1.14)

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

which with (1.13) leads to

$$E(\mathcal{X}) = JE\{\xi(\omega)\} \qquad \text{and} \qquad \text{sd}(\mathcal{X}) = \underbrace{J\sqrt{\text{var}\{\xi(\omega)\} + E\{\sigma^2(\omega)\}/J}}_{\textit{common } \omega}, \tag{1.16}$$

and standard deviation is of the same order of magnitude $J$ as the expectation itself provided $\text{var}\{\xi(\omega)\} > 0$. *Such risk can not be diversified away* by increasing the portfolio size. Note the difference from the situation without common risk factors. Standard deviation is then proportional to the *square root* $\sqrt{J}$, and becomes insignificant in comparision with the mean as $J$ grows.

The argument also works when expectation and standard deviation $\xi_j(\omega)$ and $\sigma_j(\omega)$ depend on $j$. Now their *means* $\bar{\xi}(\omega)$ and $\bar{\sigma}(\omega)$ replace $\xi(\omega)$ and $\sigma(\omega)$ in (1.16).

**Random variation in claim frequency.**

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. Consider the following two sampling regimes:

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

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

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

If all $\mu_j = \mu$, this yields.

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

Let $\xi_\mu$ and $\sigma_\mu$ be the mean and standard deviation of $\mu$, and it follows from (1.16) that

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

Note that the standard deviation is (almost) proportional to the number of policies $J$. This explains the simulated patterns in Figure 6.1 where relative random uncertainty seemed unaffected by $J$.

Things are radically different when $\mu_1, \ldots, \mu_J$ are drawn independently of each other. Mean and standard deviation for individual $N_j$ are then obtained then by inserting $J = 1$ in (1.17). This yields

$$E(N_j) = T\xi_\mu \qquad \text{and} \qquad \text{sd}(N_j) = T\sqrt{\sigma_\mu^2 + \xi_\mu/T}.$$

Since $N_1, \ldots, N_J$ are independent, both means and variances may be added to determine what they are at portfolio level. This yields

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

The mean is the same as in (1.17), but the standard deviation is changed to the familiar form proportional to $\sqrt{J}$.

**Portfolio risk and variation in claim intensity**
The eventual target is the portfolio liability $\mathcal{X}$ itself. We shall now analyse how it depends on random variation in claim intensity by inserting the expressions for $E(\mathcal{N})$ and $\text{sd}(\mathcal{N})$ into (1.15) For the mean this yields

$$E(\mathcal{X}) = J\xi_\mu\xi_z, \tag{1.19}$$

the same whether $\mu$ 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

$$\text{sd}(\mathcal{X}) \;=\; \underbrace{\sqrt{J\xi_\mu(\sigma_z^2 + \xi_z^2)}}_{\textit{for pure Poisson}} \;\times\; \underbrace{\sqrt{1 + \delta\gamma}}_{\textit{due to random } \mu} \tag{1.20}$$

where

$$\delta = \frac{\sigma_z^2}{\sigma_z^2 + \xi_z^2} \times \frac{\sigma_\mu^2}{\xi_\mu} \qquad \text{and} \qquad \gamma = \begin{cases} 1 & \text{for indvidual } \mu \\ J & \text{for common } \mu \end{cases} \tag{1.21}$$

This lengty expression tells a lot. The factor $\sqrt{1 + \delta\gamma}$ on the very right in (1.20) is caused by the uncertainty in $\mu$ and makes portfolio uncertainty grow.

But by how much? In practice $\delta$ is quite small (see Exercise 6.3.2) which leads to the following observations. Suppose $\mu_1, \dots, \mu_J$ are drawn independently of each other. Then

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

*not* a high increase in risk. On portfolio level hetereogenity between policies often contributes little extra risk. This changes drastically when $\mu$ is a collective risk factor. Now $\sqrt{1+\delta\gamma} = \sqrt{1+\delta}$ which for large $J$ could be huge.

## 1.4 The role of the conditional mean

**Introduction**

The conditional viewpoint is the topic of this section too, but now $\mathbf{X}$ is going to be a quantity *observed.* This opens for

$$\hat{Y} = \xi(\mathbf{X}) \qquad = \qquad E(Y|\mathbf{X}) \tag{1.22}$$

as a way of guessing the value of an unknown, possibly future $Y$. In practice many quantities could be hidden in $\mathbf{X}$, and bold face notation has therefore been used. We shall see below that in theory (1.22) is the best way such **prediction** of $Y$ can be carried out. That is a celebrated result in engineering, statistics and elsewhere, yet not *that* prominent in actuarial science, since we are often more concerned with summaries such as mean and percentiles than with predicting the actual outcome of a future $Y$.

But the conditional mean $E(Y|\mathbf{x})$ has another, important usage. Suppose $\mathbf{X}$ is the information possessed. The conditional mean then conveys what is expected of $Y$ given that knowledge and could be a natural *price* for a risk $Y$. Shouldn't what we charge reflect what we know, say as a sort of *conditional* pure premium? This viewpoint is of relevance in insurance and lead in finance to the theoretical interest rate curve, the most important example of this section. A word on the meaning of $\mathbf{X}$ in the present context is needed. We might think of it as all present and past observations. Theoretical literature in mathematical finance often refers to $\mathbf{X}$ as a **sigma-field** (typically denoted $\mathcal{F}$), but it is perfectly possible to understand the ideas involved without this formalism.

**Optimal prediction**

The central mathematical properties of the conditional mean are

$$\underbrace{E(\hat{Y} - Y) = 0}_{\textit{expected error}} \qquad \text{and} \qquad \underbrace{E(\hat{Y} - \tilde{Y})^2 \leq E(\tilde{Y} - Y)^2}_{\textit{expected squared error}} \quad \text{for all } \tilde{Y} = \tilde{Y}(\mathbf{X}). \tag{1.23}$$

On the right $\tilde{Y} = \tilde{Y}(\mathbf{X})$ is an arbitrary function of $\mathbf{X}$. Here the left hand side, which is merely a rephrasal of the rule of double expectation (1.13), signifies that expected prediction error $\hat{Y} - Y$ is zero. The prediction $\hat{Y}$ is thus **unbiased**; more on that concept in Chapter 7. On the right the inequality shows that the expected *squared* error is *smaller* than for any other way of utilizing the information $\mathbf{X}$. The proof is simple, and is given in Section 6.8. In this sense the conditional mean is the most accurate way information can be exploited.

**The conditional mean in property insurance**

The principal application of the conditional mean in property insurance is to *differentiated pricing*, based on individual assesment of risk. There are two traditions. One works from the record of each policy holder. Let $X_1, \ldots, X_n$ be annual claims $n$ years back. The question is what this conveys about the risk of the individual and the expected, *future* claim $\pi = E(X)$. A natural estimate is the conditional expectation given the claim record, i.e.

$$\hat{\pi} = E(X|X_1, \ldots X_n).$$

This idea will lead to the **credibility** theory in Chapter 10. The underlying model is of the common factor type where $X_1, \ldots, X_n, X$ are conditionally independent (and identically distributed) given some underlying random quantitiy $\omega$ assigend the individual.

Another (and increasingly popular) way it to link expected claim or pure premium to explanatory variables. Typical examples are age and sex of car owners. We then let experience with *the group* influence the premium charged. Usually claim *frequency* varies much stronger among the population than claim *size*, and their modelling is often carried out separately. For example, we shall work with

$$E(N|x_1, x_2, \ldots);$$

i.e. expected claim frequency given information $x_1, x_2, \ldots$ on age, sex and other things. How such models are built is discussed in Chapter 8, see also Ecercise 6.4.5

**Interest rate prediction**

The rest of the section deals with interest rates, and we shall first examine statistical forecasts. As an example consider the Vasicěk model of Section 5.7, under which the rate of interest at time $t_k$ can 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 random disturbances of the future. Suppose they are independent with zero mean. That is the standard assumption which yields

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

These formulae were derived in Section 5.7. On the left is the best prediction of $r_k$ if the Vasicěk model is true.

What would the accuarcy 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. Forecasting interest levels through purely statistical procedures is futile.

**Theoretical interest rate curves**

Conditional arguments have in this context another (and more important) usage. Zero-coupon bonds were introduced in Section 1.4. For the delevery of one money unit at $t_k = kh$ the market charges $P(0{:}k)$ today. Theoretical models for such instruments will be needed in Chapter 15. We shall write those $P(r_0, t_k)$ highlighting their link to the present rate of interest $r_0$ and the time to expiry of the bond. With the possession of such models we may simulate *future* bond prices. Simply replace $r_0$ by a simulation $r_i^*$ of its value at $t_i$ and interprete $P(r_i^*, t_k)$ as the bond prices at that time; for details see Section 15.3. Of course, as $k$ is varied $P(r_0, t_k)$ should match the observed prices $P(0{:}k)$. That is how the parameters in their expression usually are determined.

The most common construction is to take

$$P(r_0, t_k) = E_Q(D_k|r_0) \qquad \text{where} \qquad D_k = \frac{1}{1+r_1} \times \cdots \times \frac{1}{1+r_k} \tag{1.24}$$

where $D_K$ is the *stochastic discount* we might have used at $t_0 = 0$ had the future rates $r_1, \ldots, r_k$ been known. Its expected value is the theoretical value of the bond. An implicit assumption is that interest rates behave **Markovian**, see Section 6.5 below. Otherwise historical rates prior to the current one $r_0$ would influence our belief in their future values. The subscript $Q$ in (1.24) refers to risk-neutrality; see Section 3.6 and (above all) Chapter 14.

As $k$ is varied, the **term structure** of theoretical bond prices is defined. By passing to continuous time it has for many interest rate models been possible to derive simple mathematical expressions. Many of them have actually been proposed for that very purpose; see Section 6.9. Consider the Vasicěk model

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

where $h$ is included in the mathematical notation; see (**??**) and where parameters are $q$-subscripted to link them to the risk-neutral model. Calculations of (1.24) under this model were carried out in Exercises 5.7.12-16 though a standard limiting process for which $h \to 0$, $k \to \infty$ while $t = hk$ is kept fixed. That lead to the expression

$$P(r_0, t) = e^{A(t) - B(t) r_0} \tag{1.25}$$

where

$$B(t) = \frac{1 - e^{-a_q t}}{a_q} \qquad \text{and} \qquad A(t) = (B(t) - t)\left(\xi_q - \frac{\sigma_q^2}{2a_q^2}\right) - \frac{\sigma_q^2 B(t)^2}{4a_q}. \tag{1.26}$$

We may interprete $P(r_0, t)$ as the price in a Vasicěk world of a zero-coupon bond maturing at time $t$ when the initial rate of interest is $r_0$.

**Term structure by Monte Carlo**
It is perfectly feasible to approximate $P(r_0, t_k)$ through Monte Carlo when mathematical expressions are not available. The following implementation is adapted to the Black-Karisinsky model, but it can easily be modified.

**Algorithm 6.1 The Black-Karisinsky term structure**
0 Input: $m$, $\xi_q$, $a_q$, $\sigma_q$, $r_0$, $h$ 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 $P^*(k)$ to $P(r_0, t_k)$ for $k = 1, \ldots, K$. Re-runs for many different $r_0$ yield a table in $k$ and $r_0$ that could be used with evaluations as those in Section 15.

If you want the computations to run a finely meshed time scale, you must adapt the parameters as explained in Section 5.7. The examples in Figure 6.3 have been run om a crude annual one with parameters

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

where the volatilities $\sigma$ have been adjusted to make standard deviations of $r_k$ coincide at steady state. It is more illuminating to consider the yield curve than the bond prices themselves. The quantities plotted are thus

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

which is the average rate of interest over the period in question; see Section 1.4. The initial rate was varied between $r_0 = 2\%$, $4\%$ $6\%$ $8\%$ and $10\%$. All yield curves have an exponential shape up or down depending on the start. It takes in either case a long time to approach the average $\xi = 4\%$. the speed depends on of $a_q$.

## 1.5 Joint probability models

**Introduction**

A general probabilitic description of dependent random variables $X_1, \ldots, X_n$ is 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

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

under variation of $x_1, \ldots, x_n$ and $f(x_1, \ldots, x_n)$ is their $n$-fold partial derivative with respect to $x_1, \ldots, x_n$. In practice we may think of is as the likelihood of the event[4]

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

[4]Formally, in a strict mathematical sense, such events have probability zero.

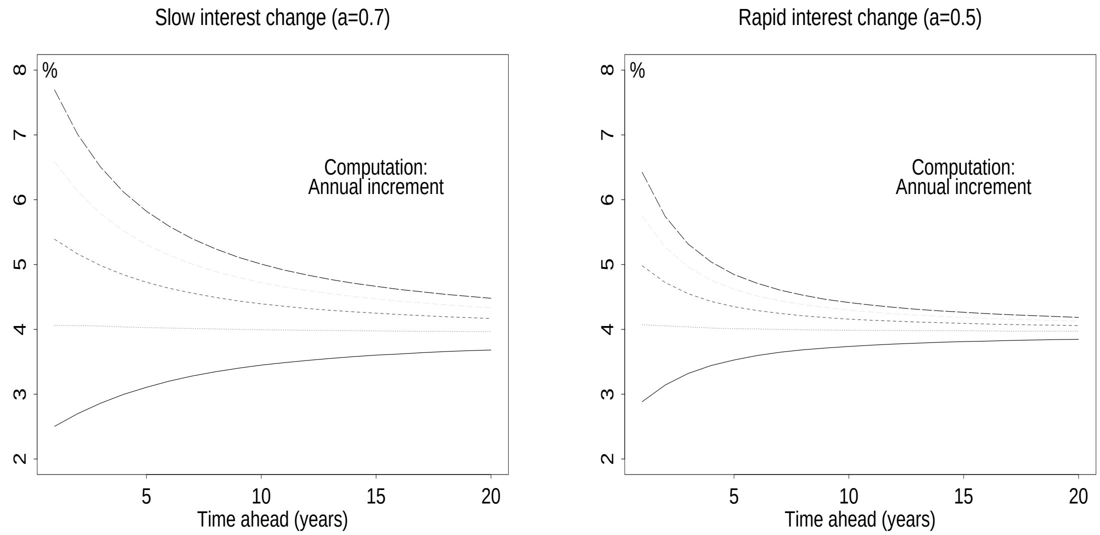



*Figure 6.2 Interest rate curves under the Black-Karisinsky model when the initial rate of interest is varied.* 10000 *simulations.*

Textbooks in probability and statistics often *start* with density functions. They 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 densities are also needed for the **likelihood** criterion in the next chapter, which often opens for the best possible use of historical data. The multinomial distribution below and the copulas in Section 6.7 are examples of modelling joint densities directly. Usually this book has followed the tactics of defining models in the way they are simulated, and we shall introduce joint density functions in that way too.

**Factorization of joint densities**

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

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

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.

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

$$f(x_1, \dots, x_n) = \underset{\textit{general factorization}}{f(x_1)f(x_2|x_1)\cdots f(x_n|x_1, \dots, x_{n-1})}, \tag{1.27}$$

which reflects that the sampling scheme above produces a Monte Carlo simulation from $f(x_1, \dots, x_n)$. In (1.27) 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

$$f(x_1,\ldots,x_n) = f(x_1)f(x_2|x_1)\cdots f(x_n|x_1). \tag{1.28}$$
*Common factor: First variable*

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

$$f(x_1,\ldots,x_n) = f(x_1)f(x_2)\cdots f(x_n). \tag{1.29}$$
*Independence*

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

$$f(x_1,\ldots,x_n) = f(x_1)f(x_2|x_1)\cdots f(x_n|x_{n-1}), \tag{1.30}$$
*Markov dependence*

where $X_k$ only depends on the preceding $X_{k-1}$, those before $t_{k-1}$ being irrelevant. This is the standard model in life insurance, and it will be discussed in the next section. Both the random walk model and the first order autoregressive model in Section 5.7 were of the Markov type. 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 $X_1^*,\ldots,X_n^*$

Examples are given in Section 6.6 and in Exercise 6.5.1.

**The general normal density**
Most famous of all joint density functions is the **Gassian one**; i.e.

$$f(\mathbf{x}) = (|2\pi\Sigma|)^{-1/2}\exp\{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\xi})'\Sigma^{-1}(\mathbf{x}-\boldsymbol{\xi})\} \tag{1.31}$$

where $\boldsymbol{\xi} = (\xi_1,\ldots,\xi_n)'$ is the vector of expectations and $\Sigma$ the covariance matrix; see Chapter 5[5]. In the present book this expression is of little importance.

**The multinomial density and delayed claims**
In some branches of insurance there are long delays between the accident that caused the damage

[5] The notation $|2\pi\Sigma|$ signifies the determinant of the matrix $2\pi\Sigma$; see Appendix B.

and its reporting and settling financially. It could take up to a couple of decades and in extreme cases even longer; the actual number of claims may not be known until long after[6]. The insurance industry is in any case responsible, and must be able to project the economic consequences of the delay; more on that in chapter 8.

Let $n$ be the number of claims arising in a given year with $N_k$ being those settled $k$ years after the event, $k = 0, 1, \ldots, K$. Here $K$ is the maximum delay and

$$N_0 + N_1 + \ldots + N_K = n,$$

Suppose we take the position that each insurance incident even *after* the event is a chance experiment in terms of how long it takes for the claim to be to liquidated. Let $q_k$ be the probability of $k$ years. Clearly

$$q_0 + \ldots + q_K = 1.$$

There are $n$ such trials, reasonably regarded as independent of each other, and elementary courses in probability show that $N_k$ must be a binomial random variable with sucess probablity $q_k$. The extension to many events simultaneously is the **multinomial** model, under which the density function is

$$f(n_0, \ldots n_K|n) = \frac{n!}{n_0!\ldots n_K!}\, q_0^{n_0} \ldots q_K^{n_K}, \qquad \text{where} \qquad n_0 + \ldots + n_K = n; \tag{1.32}$$

see Exercise 6.5.8 for its derivation. Note the conditional statement given $N = n$, which is itself a random variable (and which will not be fully known until later). Sampling can be carried out by means of the method of guide tables in Chapter 4; see also Exercise 6.5.7.

## 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, it was in Chapter 3 demonstrated that uncertainty due to life cycle movements rarely is very important. But that doesn't mean that the underlying stochastic model

[6]Injury in automobile accidents is an example. It may take long before the symptoms of a neck or back ailment emerge.

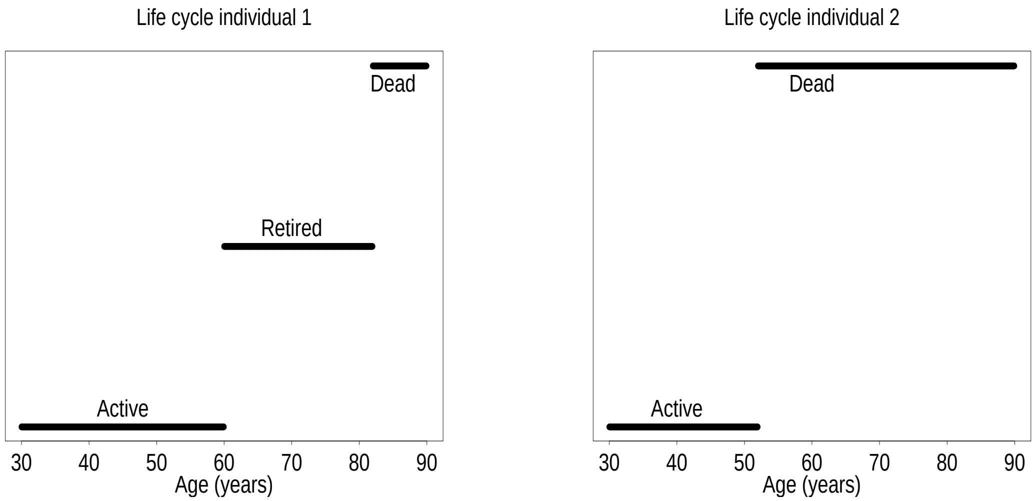



*Figure 6.3 The life cycles of two members of a pension scheme.*

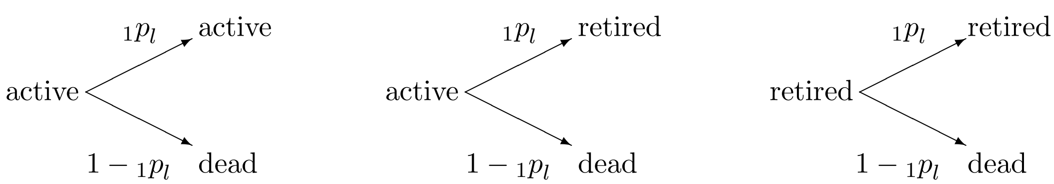
is irrelevant. It is needed both to compute the expectations defining the liabiltities *and* to eveluate portfolio uncertainty due to errors in parameters.

**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**

$$p_l(i|j) = \Pr(C_{l+1} = i|C_l = j). \tag{1.33}$$

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 indvidual. 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.

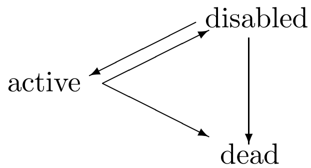
Transition probabilities are usually different for men and women (not reflected in the mathematical notation), and it is (of course) essential that they depend on age $l$. A major part of them always come from the survival probabilities ${}_1p_l$ introduced in Section 6.2; see (1.4). For a simple pension 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 (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. Both ususlly depend on age $l$ which has been suppressed in the mathematical notation.

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 | ${}_1p_l \times (1 - p_{i\|a})$ | ${}_1p_l \times p_{i\|a}$ | $1 - {}_1p_l$ | 1 |
| Disabled | ${}_1p_l \times p_{a\|i}$ | ${}_1p_l \times (1 - p_{a\|i})$ | $1 - {}_1p_l$ | 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 stays.

**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

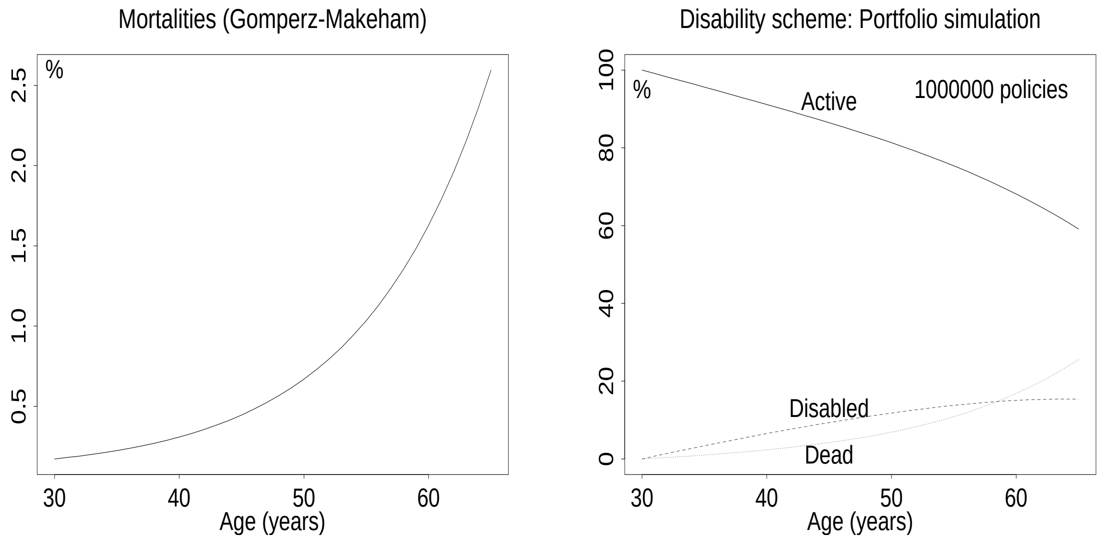



*Figure 6.4 A disability sceme in life insurance: Mortality model (left) and portfolio simulation (right).*

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

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

which is no more than speculation. Note the rehabilitation rate, which is probably too high.

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 development has been simulated using Algorithm 6.2. Details are discussed 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 true in practice is the downwards curvature in the disability curve which is due to the unrealistic, age-independent specification of the disability rate.

## 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 what coupla modelling utlizes. Dependence are formulated in terms of $U_1$ and $U_2$, which are then mapped back to $X_1$ and $X_2$ through the transformations. 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

$$H(u_1, u_2) = \Pr(U_1 \leq u_1, U_2 \leq u_2), \tag{1.34}$$

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

$$\begin{aligned} &H(u_1, 0) = 0, \qquad H(0, u_2) = 0 \\ &H(u_1, 1) = u_1, \quad\; H(1, u_2) = u_2 \end{aligned} \tag{1.35}$$

for any $u_1$ and $u_2$. For example

$$H(u_1, 1) = \Pr(U_1 \leq u_1, U_2 \leq 1) = \Pr(U_1 \leq u_1) = u_1,$$

and similar for the others. Any function $H(u_1, u_2)$ serving as a copula must satisfy (1.35).

The most immediate example is

$$H(u_1, u_2) = u_1 u_2,$$

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

$$H(u_1, u_2) = (u_1^{-\theta} + u_2^{-\theta} - 1)^{-1/\theta}, \quad 0 < u_1, u_2 < 1, \qquad (\theta > 0) \tag{1.36}$$

Here $\theta$ is a positive parameter (negative ones will be allowed later). It is easily verified that (1.35) 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 $H(u_1, u_2)$

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

$$\underbrace{F(x_1,x_2) = H(u_1,u_2)}_{\textit{copula modelling}} \qquad \text{where} \qquad \underbrace{u_1 = F_1(x_1),\;\; u_2 = F_2(x_2).}_{\textit{univariate modelling}} \tag{1.37}$$

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[7], and Sklar's result can be extended to any number of variables.

In (1.37) either of the relationships on the right may be replaced by their antitetic twin (see Section 4.5). This produces the three additional versions

$$\begin{array}{ll} u_1 = F_1(x_1),\;\; 1-u_2 = F_2(x_2) & \textit{orientation (1,2)} \\ 1-u_1 = F_1(x_1),\;\; u_2 = F_2(x_2) & \textit{orientation (2,1)} \\ 1-u_1 = F_1(x_1),\;\; 1-u_2 = F_2(x_2) & \textit{orientation (2,2),} \end{array} \tag{1.38}$$

all combined with the *same* copula on the left in (1.37) . 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 (1.36) can be extended to include *negative* $\theta$ down to $-1$, provided the mathematical expression is modified to

$$H(u_1, u_2) = \max\{(u_1^{-\theta} + u_2^{-\theta} - 1)^{-1/\theta}, 0\} \qquad (\theta \geq -1). \tag{1.39}$$

Again it is easy to check that the copula requirements (1.35) are satisfied when $\theta \geq -1$. For negative $\theta$, 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, and positive $\theta$ are more useful for actuarial science. 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 (Chapter 12). Ordinary Gaussian models can't capture this. The other plots in Figure 6.5 rotate patterns by

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

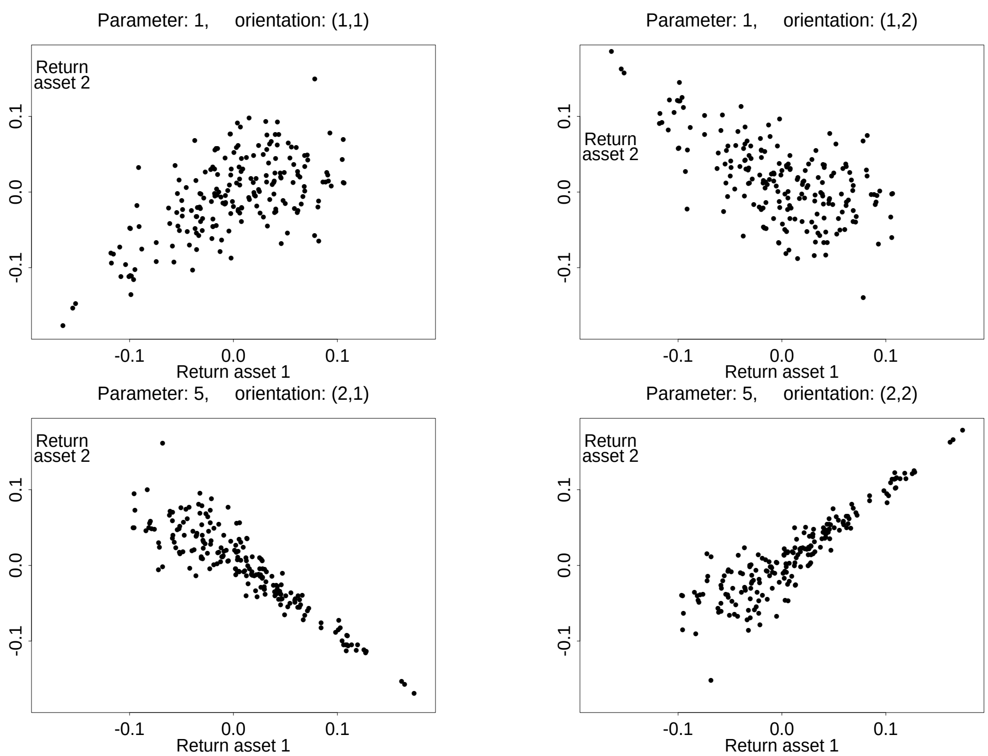



*Figure 6.5 Simulated financial returns from normals and* **Clayton** *copula.*

varying the orientation of the copula. Dependence is adjusted by moving $\theta$ (high values for strong dependence).

**Conditional distributions for copulas**

As elsewhere it is useful to examine the conditional models. Let

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

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.

$$H(u_2|u_1) = \frac{\partial H(u_1,u_2)}{\partial u_1}; \tag{1.40}$$

see Section 6.8.

For the Clayton copula (1.36) straightforward differentiaton yields

$$H(u_2|u_1) = u_1^{-(1+\theta)}(u_1^{-\theta} + u_2^{-\theta} - 1)^{-(1+1/\theta)}, \tag{1.41}$$

where

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

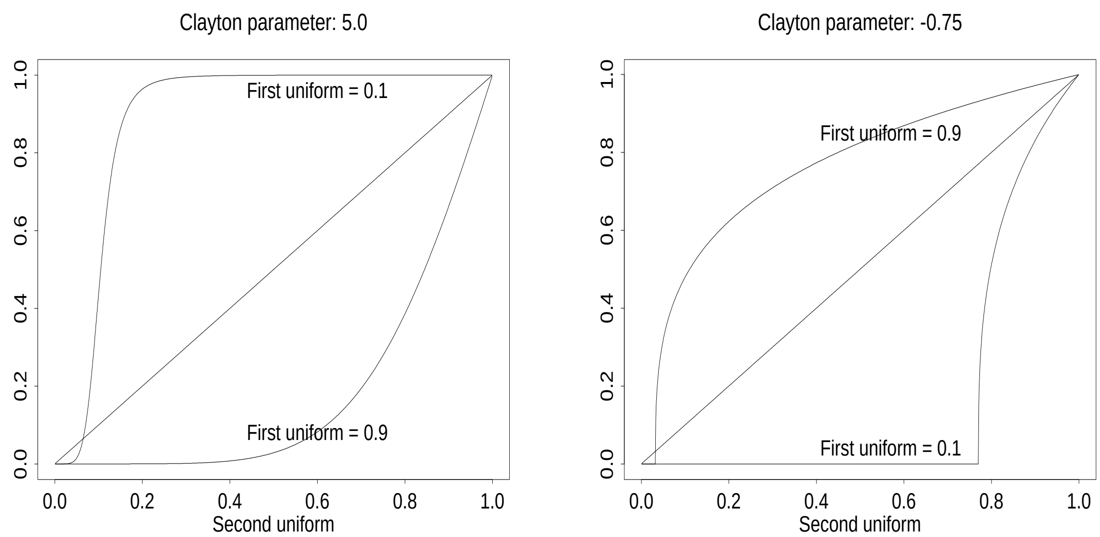



*Figure 6.6* **Conditional** *distribution functions for the* **second** *variable of a Clayton copula; given first variable marked on each curve.*

Below the lower threshold $H(u_2|u_1) = 0$. The conditional distribution functions have been plotted in Figure 6.6 for $\theta$ 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 $H(u_2|u_1)$

1 Draw $U_1^*$ and $V^* \sim$ uniform

2 Determine $U_2^*$ from

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

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 (1.41) 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

$$H(u_1, u_2) = \phi^{-1}\{\phi(u_1) + \phi(u_2)\}. \tag{1.42}$$

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

$$\phi(u) = \frac{1}{\theta}(u^{-\theta} - 1), \qquad \text{and} \qquad \phi(x)^{-1} = (1+\theta x)^{-1/\theta}, \tag{1.43}$$

where the inverse is found by solving $\phi(u) = x$ for $u$. If these expressions are inserted into (1.42), 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 (1.35) 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* $\theta$) 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 possibilites.

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 $U_2^*$

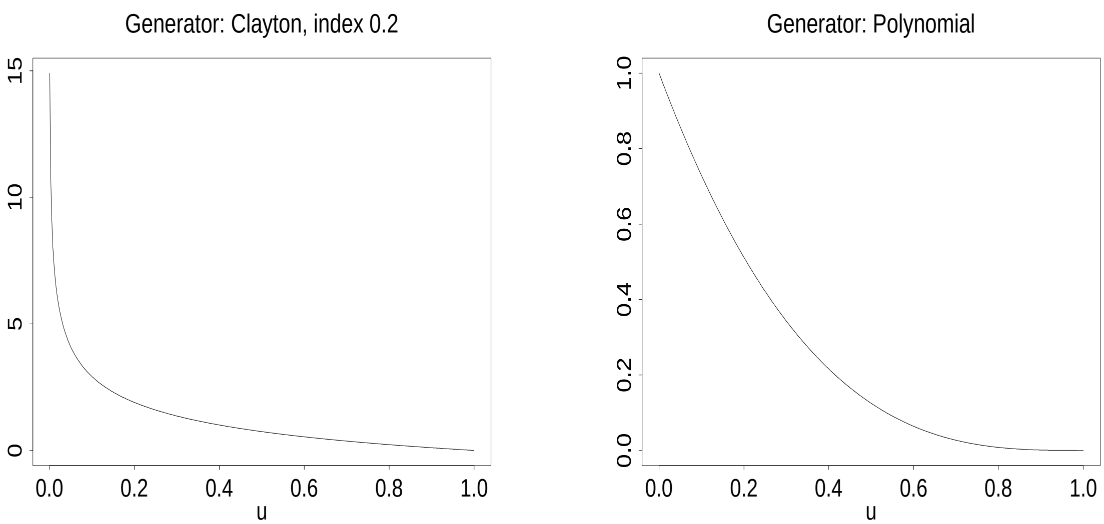



*Figure 6.7 Generator functions for Archimedean copulas*

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 $H(u_1, \ldots, u_J)$ is the joint distribution function of $J$ dependent uniform random variables $U_1, \ldots, U_J$. Mathematical conditions similar to (1.35), but more complex have to be satisfied. There is an Archimedean type which is an immediate extension of (1.42); i.e.

$$H(u_1, \ldots, u_J) = \phi^{-1}\{\phi(u_1) + \ldots + \phi(u_J)\}. \tag{1.44}$$

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

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

There are now $2^J$ ways to rotate patterns through use of antitetic 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**

Input: $\theta$
$S^* \leftarrow 0$ and $P^* \leftarrow 1$ %*Initializing auxilliary quantities*
Draw $U_1^* \sim$ uniform
For $j = 2, \ldots, J$ do
  $S^* \leftarrow S^* + (U_{j-1}^*)^{-\theta} - 1$ %*Updating from preceding uniform*
  $P^* \leftarrow P^*(U_{j-1}^*)^{1+\theta}/(1 + (j-1)\theta)$ %*Here too*

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 (1.20) 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 (1.15). 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 (1.20).

**Section 6.7.**

**Conditional distributions for copulas** Recall that

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

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

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

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

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

where $h(u_2|u_1)$ is the conditional density of $U_2$. This is (1.40).

**Justifying Algorithm 6.4** An Archimedean copula $H(u_1, u_2)$ is according to (1.42) defined through

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

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

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

so that

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

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

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

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

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

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 (1.43) are inserted into (1.44), it follows that the mathematical expression for the $J$-dimensional Clayton copula is

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

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

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

It is straightforwardly derived that

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

or

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

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

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

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 $a_{j-1}<X_{1i}^*\le a_j$ 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 $k = 1, 2, \dots, 30$.

**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 $\mu$ 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 $\sigma$ 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 $\sigma$ 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 $\mathcal{N} \sim$Poisson$(J\mu T)$. Show that the formulas (1.15) 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 $\mu$ 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 (1.19) and (1.20). **a)** Show that $\delta \le \sigma_\mu^2/\xi_\mu$. **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 $\sqrt{1+\delta} \le 1.023$ under the specification in b) and argue that the *added* portfolio risk due to the hetereogenity in $\mu$ accounts for no more than 2.3% of the total value of 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 $\mu$ varies randomly, but now as a common parameter for all policy holders. **a)** Go back to (1.19) 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 $\mu$-*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 $\sigma_z/\xi_z$. Is the impact of $\mu$-variability larger or smaller for heavy-tailed claim size distributions than for lighter ones?

**Exercise 6.3.4** Suppose that $X_1, \dots, X_J$ are conditionally independent and identically distributed given a common factor $\omega$. **a)** Explain that (1.16) 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 sd$(\omega)$ are the conditional mean and standard deviation. **b)** Show that

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

**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 $\mu$ 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

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

**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 $\xi$ and $\sigma$ 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 $\varepsilon_1$, $\varepsilon_2$ and $Z$ are independent and the two former are $N(0, 1)$ with correlation $\rho$.

**Section 6.4**

**Exercise 6.4.1** Let $X_1$ and $X_2$ be dependent normal variables with expectations $\xi_1$ and $\xi_2$, standard deviations $\sigma_1$ and $\sigma_2$ and correlation $\rho$. **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 $\rho$ should from this viewpoint be interpreted through $\rho^2$, as claimed in Section 5.2.

**Exercise 6.4.2** Claim intensities $\mu$ 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 $\mu$ 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 $\mu_0$ and $\beta_0$ are parameters. **a)** What is the meraning of the parameters $\mu_0$ and $\beta$? **b)** Determine them so that $\mu = 10\%$ at age 18 and 5% at age 60 and plot the relationship between $x$ and $\mu$. 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 $\varepsilon_1, \varepsilon_2, \dots$ 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 $X_1, \dots, X_J$ are conditionally normal given $Z = z$ with expectations $\xi_i$ 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 $\xi_1, \xi_2, \xi_3$, volatilities $\sigma_{01} sqrtz, \sigma_{02}\sqrt{z}, \sigma_{03}\sqrt{z}$ and correlations $\rho_{ij}$. **a)** Explain how the log-returns are samples. **b)** Carry out the sampling 1000 times when

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

**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**[8] 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 $\xi$ 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 (1.27). 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 $N_K^* \leftarrow n - S^*$.

[8]ARCH stands for **a**uto**r**egressive, **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 (1.32) 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 $\theta$ 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 $\theta$ 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 $\theta$ 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 $\zeta$ (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

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

**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

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

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

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

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 (1.40).]. **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 (1.39) 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 (1.37) 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 (1.35). **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 (1.35) 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 (1.36) with positive $\theta$ with generator $\phi(u) = (u^{-\theta} - 1)/\theta$. Show that the key part of Algorithm 6.4 (lines 2 and 3) is solved by

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