STOCHASTIC PROCESSES: LEARNING THE LANGUAGE

By A. J. G. Cairns, D. C. M. Dickson, A. S. Macdonald,
H. R. Waters and M. Willder

ABSTRACT

Stochastic processes are becoming more important to actuaries: they underlie much of modern finance, mortality analysis and general insurance; and they are reappearing in the actuarial syllabus. They are immensely useful, not because they lead to more advanced mathematics (though they can do that) but because they form the common language of workers in many areas that overlap actuarial science. It is precisely because most financial and insurance risks involve events unfolding as time passes that models based on processes turn out to be most natural. This paper is an introduction to the language of stochastic processes. We do not give rigorous definitions or derivations; our purpose is to introduce the vocabulary, and then survey some applications in life insurance, finance and general insurance.

KEYWORDS

Financial Mathematics; General Insurance Mathematics; Life Insurance Mathematics; Stochastic Processes

AUTHORS' ADDRESSES

A. J. G. Cairns, M.A., Ph.D., F.F.A., Department of Actuarial Mathematics and Statistics, Heriot-Watt University, Edinburgh EH14 4AS, U.K. Tel: +44(0)131-451-3245; Fax: +44(0)131-451-3249; E-mail: A.J.G.Cairns@ma.hw.ac.uk

D. C. M. Dickson, B.Sc., Ph.D., F.F.A., F.I.A.A., Centre for Actuarial Studies, The University of Melbourne, Parkville, Victoria 3052, Australia. Tel: +61(0)3-9344-4727; Fax: +61-3-9344-6899; E-mail: ddickson@cupid.ecom.unimelb.edu.au

A. S. Macdonald, B.Sc., Ph.D., F.F.A., Department of Actuarial Mathematics and Statistics, Heriot-Watt University, Edinburgh EH14 4AS, U.K. Tel: +44(0)131-451-3209; Fax: +44(0)131-451-3249; E-mail: A.S.Macdonald@ma.hw.ac.uk

H. R. Waters, M.A., D.Phil., F.I.A., F.F.A., Department of Actuarial Mathematics and Statistics, Heriot-Watt University, Edinburgh EH14 4AS, U.K. Tel: +44(0)131-451-3211; Fax: +44(0) 131-451-3249; E-mail: H.R.Waters@ma.hw.ac.uk

M. Willder, B.A., F.I.A., Department of Actuarial Mathematics and Statistics, Heriot-Watt University, Edinburgh EH14 4AS, U.K. Tel: +44(0)131-451-3204; Fax: +44(0)131-451-3249; E-mail: M.Willder@ma.hw.ac.uk

1. Introduction

April 2000 will see the introduction of the new 100 series examinations. Among these is a new subject called Stochastic Modelling (103). In this paper we will introduce you to the main concepts in stochastic modelling, and then illustrate their application to Life Insurance Mathematics, Finance and General Insurance. We hope that this paper will encourage you to find out more about stochastic processes and how you can use them in your work.

This paper should be of particular interest to qualified actuaries and trainees who have already completed subjects A to D and so (to their deep regret) will miss the opportunity to sit the new examination. This paper assumes that you are already familiar with the basics of probability and the actuarial applications.

2. Familiar Territory

We will start off in what we hope is familiar territory. Consider two simple experiments:
(a) spinning a fair coin; or
(b) rolling an ordinary six sided die.

Each of these experiments has a number of possible **outcomes**:
(a) the possible outcomes are {H} (heads) and {T} (tails); and
(b) the possible outcomes are {1}, {2}, {3}, {4}, {5} and {6}.

Each of these outcomes has a **probability** associated with it:
(a) $\mathrm{P}[H] = 0.5 = \mathrm{P}[T]$; and
(b) $\mathrm{P}[1] = \frac{1}{6} = \ldots = \mathrm{P}[6]$.

The set of possible outcomes from experiment 1 is rather limited compared to that from experiment 2. For experiment 2 we can consider more complicated **events**, each of which is just a subset of the set of all possible outcomes. For example, we could consider the event {even number}, which is equivalent to {2, 4 or 6}, or the event {less than or equal to 4}, which is equivalent to {1, 2, 3 or 4}. Probabilities for these events are calculated by summing the probabilities of the corresponding individual outcomes, so that:

$$\mathrm{P}[\text{even number}] = \mathrm{P}[2] + \mathrm{P}[4] + \mathrm{P}[6] = 3 \times \frac{1}{6} = \frac{1}{2}$$

A real valued **random variable** is a function which associates a real number with each possible outcome from an experiment. For example, for the coin spinning experiment we could define the random variable X to be 1 if the outcome is {H} and 0 if the outcome is {T}.

Now suppose our experiment is to spin our coin 100 times. We now have 2^{100} possible outcomes. We can define events such as {the first spin gives Heads and the second spin gives Heads} and, using the presumed **independence** of the results of different spins, we can calculate the probability of this event as $\frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$.

Consider the random variable X_n, for $n = 1, 2, \ldots, 100$ which is defined to be the number of Heads in the first n spins of the coin. Probabilities for X_n come from the binomial distribution, so that:

$$\mathrm{P}[X_n = m] = \binom{n}{m} \times \left(\frac{1}{2}\right)^m \times \left(\frac{1}{2}\right)^{n-m} \qquad \text{for } m = 0, 1, \ldots, n.$$

We can also consider **conditional probabilities** for X_{n+k} **given** the value of X_n. For example:

$$\begin{aligned}
\mathrm{P}[X_{37} &= 20 \mid X_{36} = 19] = \frac{1}{2} \\
\mathrm{P}[X_{37} &= 19 \mid X_{36} = 19] = \frac{1}{2} \\
\mathrm{P}[X_{37} &= m \mid X_{36} = 19] = 0 \quad \text{if } m \neq 19, 20.
\end{aligned}$$

From these probabilities we can calculate the **conditional expectation** of X_{37} given that $X_{36} = 19$. This is written $\mathrm{E}[X_{37} \mid X_{36} = 19]$ and its value is 19.5. If we had not specified the value of X_{36}, then we could still say that $\mathrm{E}[X_{37} \mid X_{36}] = X_{36} + 0.5$. There are two points to note here:

(a) $\mathrm{E}[X_{37} \mid X_{36}]$ denotes the expected value of X_{37} given some **information** about what happened in the first 36 spins of the coin; and
(b) $\mathrm{E}[X_{37} \mid X_{36}]$ is itself a random variable whose value is determined by the value taken by X_{36}. In other words, $\mathrm{E}[X_{37} \mid X_{36}]$ is a function of X_{36}.

The language of elementary probability theory has been adequate for describing the ideas introduced in this section. However, when we consider more complex situations, we will need a more precise language. This more precise language will be introduced in the next section, although in some places we will sacrifice mathematical rigour for brevity and clarity.

3. Basic Terminology

3.1 *Uncertainty*

In this paper we are concerned with uncertainty. We will be interested in the results of 'experiments' that cannot be fully predicted beforehand. These experiments could be as varied as measuring the speed of a car, recording the result on a die, or calculating the size of reserves.

3.2 *Probability Triples*

We will introduce the mathematical shorthand $(\Omega, \mathcal{F}, \mathrm{P})$, known as a **probability triple**. The three parts of $(\Omega, \mathcal{F}, \mathrm{P})$ are the answers to three very important questions when dealing with uncertain experiments, namely:
(a) What are the possible outcomes of an experiment?
(b) What information do we have about the outcome of an experiment?
(c) What is the underlying probability of each outcome occurring?

We will start by explaining the use and meaning of the terminology $(\Omega, \mathcal{F}, \mathrm{P})$.

3.3 *Sample Spaces*

The **sample space** Ω is the set of all the possible outcomes, ω, of the experiment. We call each outcome a **sample point**. In an example of rolling a 6 sided die our sample space is simply:

$$\Omega = \{1, 2, 3, 4, 5, 6\}$$

We see that in this case we have 6 sample points. We would now express the outcome of a 4 being rolled as $\omega = 4$.

An **event** is a subset of the sample space. In our example we would represent the event of an odd number being rolled by the subset $\{1, 3, 5\}$.

3.4 σ-*algebras*

We denote by $\mathcal{F}$ the set of all **events** in which we could possibly be interested. To make the mathematics work, we insist that $\mathcal{F}$ contains the empty set $\emptyset$, the whole sample space Ω, and all (countable[1]) unions, intersections and complements of its members. With these conditions, $\mathcal{F}$ is called a **σ-algebra** of events.

[1] A countable infinite set is one which can be put into one-to-one correspondence with the positive integers. The set of all rational numbers is countable, but the set of all real numbers is not. All finite sets are countable.

A **sub-σ-algebra** of $\mathcal{F}$ is a subset $\mathcal{G} \subseteq \mathcal{F}$ which satisfies the same conditions as $\mathcal{F}$; that is, $\mathcal{G}$ contains $\emptyset$, the whole sample space Ω, and all (countable) unions, intersections and complements of its members. For example, for the die-rolling experiment, we can take $\mathcal{F}$ to be the set of all subsets of $\Omega = \{1,2,3,4,5,6\}$; then:

$$\mathcal{G}_1 = \{\emptyset, \Omega, \{1,2,3,4\}, \{5,6\}\}$$

is a sub-σ-algebra, but:

$$\mathcal{G}_2 = \{\emptyset, \Omega, \{1,2,3,4\}, \{6\}\}$$

is not, since the complement of the set $\{6\}$ does not belong to $\mathcal{G}_2$.

3.5 *Random Variables*

A real-valued random variable, X, is a real-valued function defined on the sample space Ω.

3.6 *Probability Measure*

We now come to our third question — what is the underlying probability of an outcome occurring? To answer this we extend our usual understanding of probability distribution to the concept of probability measure. A **probability measure**, P , has the following properties:

(a) P is a mapping from $\mathcal{F}$ to the interval $[0,1]$; that is, each element of $\mathcal{F}$ is assigned a non-negative real number between 0 and 1.

(b) The probability of a countable union of disjoint members of $\mathcal{F}$ is the sum of the individual probabilities of each element; that is:

$$\mathrm{P}\left(\cup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mathrm{P}(A_i) \,\text{for}\, A_i \in \mathcal{F} \,\text{and}\, A_i \cap A_j = \emptyset, \text{for all}\, i \neq j$$

(c) $\mathrm{P}(\Omega) = 1$; that is, an outcome $\omega \in \Omega$ occurs with probability 1.

The three axioms above are consistent with our usual understanding of probability. For our die rolling experiment on the pair $(\Omega, \mathcal{F})$, we could have a very simple measure which assigns a probability of $\frac{1}{6}$ to each of the outcomes $\{1\}, \{2\}, \{3\}, \{4\}, \{5\}$, and $\{6\}$.

Now consider a biased die where the probability of an odd number is twice that of an even number. We now need a new measure P^* where $\mathrm{P}^*(\{1\}) = \mathrm{P}^*(\{3\}) = \mathrm{P}^*(\{5\}) = \frac{2}{9}$ and $\mathrm{P}^*(\{2\}) = \mathrm{P}^*(\{4\}) = \mathrm{P}^*(\{6\}) = \frac{1}{9}$. We see that this new measure P^* still satisfies the axioms above, but note that the sample space Ω and the σ-algebra $\mathcal{F}$ are unchanged. So we have shown that it is possible to define two different probability measures on the same sample space and σ-algebra, namely $(\Omega, \mathcal{F}, \mathrm{P})$ and $(\Omega, \mathcal{F}, \mathrm{P}^*)$.

3.7 *Stochastic Processes*

A **stochastic process** is a collection of random variables indexed by time; $\{X_n\}_{n=1}^{\infty}$ is a **discrete time** stochastic process, and $\{X_t\}_{t \geq 0}$ is a **continuous time** stochastic process. Stochastic processes are useful for modelling situations where, at any given time, the value of some quantity is uncertain, for example the price of a share, and we want to study the development of this quantity over time. An example of a stochastic process $\{X_n\}_{n=1}^{\infty}$ was given in Section 2, where X_n was the number of heads in the first n spins of a coin.

A **sample path** for a stochastic process $\{X_t, t \in \mathcal{T}\}$ ordered by some time set $\mathcal{T}$, is the realised set of random variables $\{X_t(\omega), t \in \mathcal{T}\}$ for an outcome $\omega \in \Omega$.

3.8 *Information*

Consider the stochastic process $\{X_n\}_{n=1}^{\infty}$ introduced in Section 2. As discussed there, the conditional expectation $\mathrm{E}[X_{n+m}|X_n]$ is a random variable which depends on the value taken by X_n. Because of the nature of this particular stochastic process, the value of $\mathrm{E}[X_{n+m}|X_n]$ is the same as $\mathrm{E}[X_{n+m}|\{X_k\}_{k=1}^{n}]$. In other words, knowing the values of $X_1, X_2, \ldots, X_{n-1}$ does not change the value of $\mathrm{E}[X_{n+m}|X_n]$. Now let $\mathcal{F}_n$ be the sub-σ-algebra created from all the possible events, together with their possible unions, intersections and complements, that could have happened in the first n spins of the coin. Then $\mathcal{F}_n$ represents the **information** we would have after n spins, from knowing the values of $X_1, X_2, \ldots, X_n$. In this case, we would describe $\mathcal{F}_n$ as the sub-σ-algebra **generated** by $X_1, X_2, \ldots, X_n$, and write $\mathcal{F}_n = \sigma(X_1, X_2, \ldots, X_n)$. The conditional expectation $\mathrm{E}[X_{n+m}|\{X_k\}_{k=1}^{n}]$ could be written $\mathrm{E}[X_{n+m}|\mathcal{F}_n]$.

More generally, our information at time t is a σ-algebra $\mathcal{F}_t$ containing those events which, at time t, we would know either had happened or had not happened.

3.9 *Filtrations*

A **filtration** is any set of σ-algebras $\{\mathcal{F}_t\}$ where $\mathcal{F}_t \subseteq \mathcal{F}_s$ for all $t < s$. So we have a sequence of increasing amounts of information where each member $\mathcal{F}_t$ contains all the information in prior members.

Usually $\mathcal{F}_t$ contains all the information revealed up to time t, that is, we do not delete any of our old information. Then at a later time, s, we have more information, $\mathcal{F}_s$, because we add to the original information the information we have obtained between times t and s.

For our coin-tossing experiment, the information provided by the filtration $\mathcal{F}_t$ should allow us to reconstruct the result of all the coin tosses up to and including time t, but not after time t. If $\mathcal{F}_t$ recorded the results of the last three tosses only, it would not lead to a filtration since $\mathcal{F}_t$ would tell us nothing about the $(t-3)^{th}$ toss.

If $\mathcal{F}_t (t \geq 0)$ is a filtration of a process X_t (taking continuous time as an example), then we have the **Tower Law of conditional expectations**. That is, for $r \leq s \leq t$:

$$\mathrm{E}[\mathrm{E}[X_t|\mathcal{F}_s]|\mathcal{F}_r] = \mathrm{E}[X_t|\mathcal{F}_r].$$

In words, suppose that at time r we want to compute $\mathrm{E}[X_t|\mathcal{F}_r]$. We could do so directly (as on the right side above) or indirectly, by conditioning on the history of the process up to some future time s (as on the left side above). The Tower Law says that we get the same answer.

3.10 *Stopping Times*

A random variable T mapping Ω to the time index set $\mathcal{T}$ is a **stopping time** if and only if:

$$\{\omega : T(\omega) = t\} \in \mathcal{F}_t \text{ for all } t \in \mathcal{T}.$$

Intuitively, a stopping time for a stochastic process is a rule for stopping this process such that the decision to stop at, say, time t can be taken only on the basis of information available at time t. For example, let X_t represent the price of a particular share at time t and consider the following two definitions:

(a) T is the first time the process $\{X_t\}$ reaches the value 120; or
(b) T is the time when the process $\{X_t\}$ reaches its maximum value.

Definition (a) defines a stopping time for the process because the decision to set $T = t$ means that the process reaches the value 120 for the first time at time t, and this information should be known at time t. Definition (b) does not define a stopping time for the process because setting $T = t$ requires knowledge of the values of the process before *and after* time t.

3.11 *Martingales*

If $\{\mathcal{F}_t\}_{t\geq 0}$ is a filtration, a **martingale with respect to** $\{\mathcal{F}_t\}_{t\geq 0}$ is a stochastic process $\{X_t\}$ with the properties that:
(a) $E(|X_t|) < \infty$ for all t;
(b) $E(X_t|\mathcal{F}_{t-1}) = X_{t-1}$, where t is a discrete index; or
(c) $E(X_t|\mathcal{F}_s) = X_s$ for all $s < t$, where t is a continuous index.

A consequence of either (b) or (c) is that:

$$\mathrm{E}[X_t] = \mathrm{E}[X_s] \quad \text{for any } t \text{ and } s.$$

A very useful property of martingales is that the expectation is unchanged if we replace t by a stopping time T for the process, so that:

$$\mathrm{E}[X_T] = \mathrm{E}[X_s] \quad \text{for any } t \text{ and } s.$$

This is the so-called **Optional Stopping Theorem**.

The word martingale has its origins in gambling games. For example take X_t to be a gambler's funds at time t. Given the information $\mathcal{F}_{t-1}$, we know the size of the gambler's funds at time $t-1$ are X_{t-1}. For a fair game (zero expected profit), the expected value of funds after a further round of the game at time t would equal X_{t-1}.

The study of martingales is a large and important field in probability. We find that many results of interest in actuarial science can be proved quickly by spotting that we have a martingale and then applying the appropriate martingale theorems.

3.12 *Markov Chains*

A **Markov chain** is a stochastic process $\{X_t\}$ where

$$\mathrm{P}(X_t = x\,|\,X_r = x_r \,\text{and}\, X_s = x_s) = \mathrm{P}(X_t = x\,|\,X_s = x_s) \quad \text{for all}\, r \leq s \leq t$$

We are interested in the probability that a stochastic process will have a certain value in the future. We may be given information as to the position of the stochastic process at certain times in the past, and this information may affect the probability of the future outcome. However for a Markov Chain the only relevant information is the most recent known value of the stochastic process. Any additional information prior to the most recent value will not change the probability.

For example, consider our die rolling experiment where N_r was the number of sixes rolled in the first r rolls. Given $N_2 = 1$, then the probability that $N_4 = 3$ is $\frac{1}{36}$ using a fair die. This probability is not altered if we also know that $N_1 = 1$.

3.13 *Further Reading*

Many textbooks cover the theory of stochastic processes. Two of the best are Grimmet & Stirzaker (1992), which starts with the basics of probability and builds up to ideas such as markov chains and martingales; and Williams (1991), which is specifically a book about martingales, but does give a very rigorous treatment of of the $(\Omega, \mathcal{F}, \mathrm{P})$ terminology.

4. Life Insurance Mathematics

4.1 *Introduction*

The aim of this section is to formulate life insurance mathematics in terms of stochastic processes. The motivation for this is the observation that life and related insurances depend on life events (death, illness and so on) that, in sequence, form an individual's life history. It is this life history that we regard as the sample path of a suitable stochastic process. The simplest such life event is death, and precisely because of its simplicity it can be modelled successfully without resorting to stochastic processes (for example, by regarding the remaining lifetime as a random variable). Other life events are not so simple, so it is more important to have regard to the life history when we try to formulate models. Hence stochastic processes form a natural starting point.

To keep matters clear, we will develop the simplest possible example, represented in an intuitive way by the two-state (or single decrement) model in Figure 1. Of course, this process — death as a single decrement — is very familiar, so at first it seems that all we do is express familiar results in not-so-familiar language. Of itself this offers nothing new, but, we emphasise, the payoff comes when we must model more complicated life histories.

(a) All the tools developed in the case of this simple process carry over to more complicated processes, such as are often needed to model illness or long term care.
(b) The useful tools turn out to be exactly those that are also needed in modern financial mathematics. In particular, **stochastic integrals** and **conditional expectation** are key ideas. So, instead of acquiring two different toolkits, one will do for both.

The main difference between financial mathematics and life insurance mathematics is that the former is based on processes with continuous paths, while the latter is based on processes with jumps[2]. The fundamental objects in life insurance mathematics are stochastic processes called 'counting processes'.

As will be obvious from the references, this section is based on the work of Professor Ragnar Norberg.

[2]It might be more accurate to say that, in financial mathematics, the easy examples are provided by continuous-path processes, and discontinuities make the mathematics much harder, while in life insurance mathematics it is the other way round. However, Norberg (1995b) suggests an interesting alternative point of view.

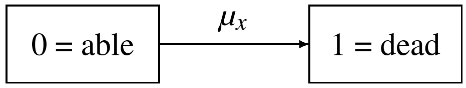

Figure 1: A two state model of mortality

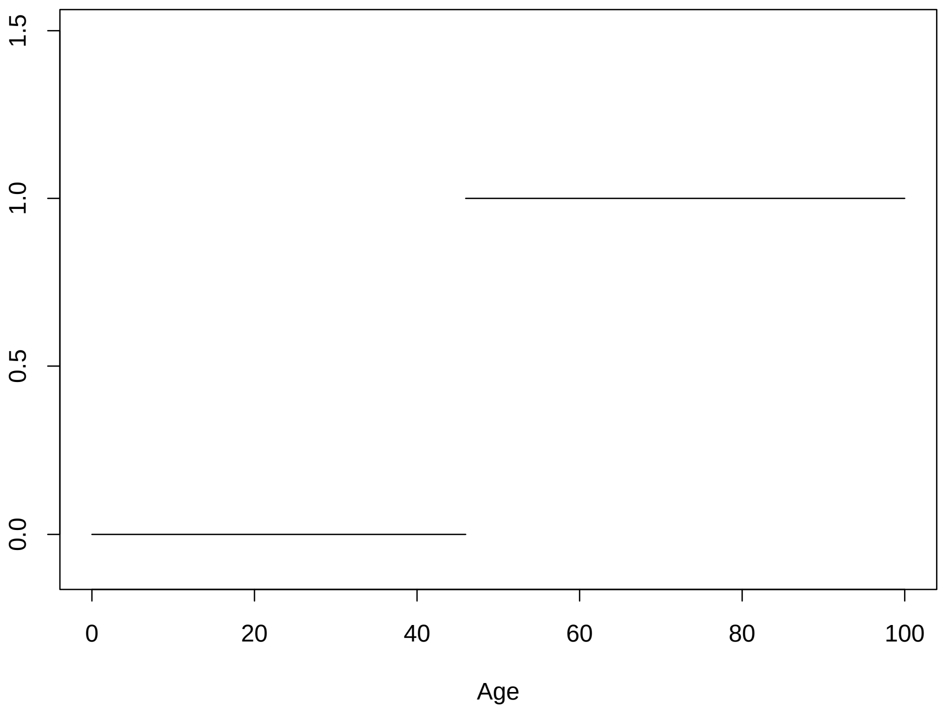

Figure 2: A sample path $N_{01}(t)$ of a counting process: death at age 46

4.2 *Counting Processes*

Figure 1 represents a two-state Markov process, with **transition intensity** ('force of mortality') μ_x depending on age x. For convenience, we assign the number 0 to the able state, and the number 1 to the dead state. A typical sample path of this process might then look like Figure 2, where a life dies at age 46. The sample path is a function of time; we call it $N_{01}(t)$. $N_{01}(t)$ indicates whether death has yet occurred[3]. Looked at another way, $N_{01}(t)$ counts the number of events that have happened up to and including time t. Do not think that because only one type of event can occur, and that only once, this 'counting' interpretation is trivial: far from it. It is what defines a counting process.

We pay close attention to the **increments** of the sample path $N_{01}(t)$. They are very simple. If the process does not jump at time t, the increment is 0. We write this as $dN_{01}(t) = 0$. If the process does jump at time t, the increment is 1. We write this as $dN_{01}(t) = 1$. (Sometimes you

[3]Strictly speaking, our sample space Ω is the space of all functions like Figure 2, beginning at 0 and jumping to 1 at some time, and the particular sample path in Figure 2 is a point $\omega \in \Omega$.

will see $\Delta N_{01}(t)$ instead of $dN_{01}(t)$; here it does not matter.)

Discrete increments like $dN_{01}(t)$ are, for counting processes, what the first derivative d/dx is for processes with differentiable sample paths. Just as a differentiable sample path can be reconstructed from its derivative (by integration) so can a counting process be reconstructed from its increments (also by integration). That leads us to the **stochastic integral**.

4.3 *The Stochastic Integral*

Begin with a discrete-time counting process, say one which can jump only at integer times. Then by definition, $dN_{01}(t) = 0$ at all non-integer times, and $dN_{01}(t) = 1$ at no more than one integer time. Can we reconstruct $N_{01}(t)$ from its increments $dN_{01}(t)$? To be specific, can we find $N_{01}(T)$? (T need not be an integer). Let $J(T)$ be the set of all possible jump times up to and including T (that is, all integers $\leq T$). Then:

$$N_{01}(T) = \sum_{t \in J(T)} dN_{01}(t). \tag{1}$$

Suppose $N_{01}(t)$ is still discrete-time, but can jump at more points: for example at the end of each month. Again, define $J(T)$ as the set of all possible jump times up to and including T, and equation (1) remains valid. This works for any discrete set of possible jump times, no matter how refined it is (years, months, days, minutes, nanoseconds ...). What happens in the limit?

(a) the counting process becomes the continuous-time version with which we started;

(b) the set of possible jump times $J(T)$ becomes the interval $(0,T]$; and

(c) the sum for $N_{01}(T)$ becomes an integral:

$$N_{01}(T) = \int_{t \in J(T)} dN_{01}(t) = \int_0^T dN_{01}(t). \tag{2}$$

The integral in equation (2) is a stochastic integral. Regarded as a function of T, it is a stochastic process[4]. This idea is very useful; it lets us write down values of assurances and annuities.

4.4 *Assurances and Annuities*

Consider a whole life assurance paying £1 at the moment of death. What is its present value at age x (call it X)? In Subjects A2 and 104, one way of writing this down is introduced: define T_x as the time until death of a life aged x (a random variable) and then the present value of the assurance is $X = v^{T_x} = e^{-\delta T_x}$ (in the usual notation).

We can also write this as a stochastic integral. The present value of £1 paid at time t is v^t. If the life does not die at time t, the increment of the counting process N is $dN_{01}(t) = 0$, and the present value of the payment is $v^t dN_{01}(t) = 0$. If the life does die at time t, the increment of N is $dN_{01}(t) = 1$, and the present value of the payment is $v^t dN_{01}(t) = v^t$. Adding up (integrating) we get:

[4]The stochastic integrals in this section are stochastic just because sample paths of the stochastic process $N_{01}(t)$ are involved in their definitions. Given the sample path of $N_{01}(t)$, these integrals are constructed in the same way as their deterministic counterparts. The stochastic integrals needed in financial mathematics, called Itô integrals, are a bit different.

$$X = \int_0^\infty v^t dN_{01}(t). \tag{3}$$

Annuities can also be written down as stochastic integrals, with a little more notation. Consider a life annuity of 1 per annum payable continuously, and let Y be its present value. Define a stochastic process $I_0(t)$ as follows: $I_0(t) = 1$ if the life is alive at time t, and $I_0(t) = 0$ otherwise. This is an **indicator process**; it takes the value 1 or 0 depending on whether or not a given status is fulfilled. Then:

$$Y = \int_0^\infty v^t I_0(t) dt. \tag{4}$$

Given the sample path, this is a perfectly ordinary integral, but since the sample path is random, so is Y. Defining $X(T)$ and $Y(T)$ as the present value of payments up to time T, we can write down the stochastic processes:

$$X(T) = \int_0^T v^t dN_{01}(t) \qquad \text{and} \qquad Y(T) = \int_0^T v^t I_0(t) dt. \tag{5}$$

4.5 *The Elements of Life Insurance Mathematics*

Guided by these examples, we can now write down the elements of life insurance mathematics in terms of counting processes. This was first done surprisingly recently (Hoem & Aalen, 1978; Ramlau-Hansen 1988; Norberg 1990, 1991). We start with payment functions:

(a) if $N = 0$ at time t (the life is alive), an annuity is payable continuously at rate $a_0(t)$ per annum; and
(b) if N jumps from 0 to 1 at time t (the life dies), a sum assured of $A_{01}(t)$ is paid.

Noting the obvious, premiums can be treated as a negative annuity, and these definitions can be extended to any multiple state model. Also without difficulty, discrete annuity or pure endowment payments can also be accommodated, but we leave them out for simplicity.

The quantities $a_0(t)$ and $A_{01}(t)$ are functions of time, but need not be stochastic processes. They define payments that will be made, depending on events, but they do not represent the events themselves. In the case of a non-profit assurance, for example, they will be deterministic functions of age. The payments actually made can be expressed as a rate, $dL(t)$:

$$dL(t) = A_{01}(t)dN_{01}(t) + a_0(t)I_0(t)dt. \tag{6}$$

This gives the net rate of payment, 'during' the time interval t to $t + dt$, depending on events. We suppose that no payments are made after time T (T could be ∞). The cumulative payment is then:

$$L = \int_0^T dL(t) = \int_0^T A_{01}(t)dN_{01}(t) + \int_0^T a_0(t)I_0(t)dt \tag{7}$$

and the value of the cumulative payment at time 0, denoted $V(0)$, is:

$$V(0) = \int_0^T v^t dL(t) = \int_0^T v^t A_{01}(t) dN_{01}(t) + \int_0^T v^t a_0(t) I_0(t) dt \tag{8}$$

This quantity is the main target of study. Compare it with equation (5); it simply allows for more general payments. It is a stochastic process, as a function of T, since it now represents the payments made depending on the particular life history (that is, the sample path of $N_{01}(t)$).

We also make use of the accumulated/discounted value of the payments at any time s, denoted $V(s)$:

$$V(s) = \frac{1}{v^s} \int_0^T v^t dL(t) = \frac{1}{v^s} \int_0^T v^t A_{01}(t) dN_{01}(t) + \frac{1}{v^s} \int_0^T v^t a_0(t) I_0(t) dt. \tag{9}$$

4.6 *Stochastic Interest Rates*

Although we have written the discount function as v^t, implicitly assuming a constant, deterministic interest rate, this is not necessary at this stage. We could just as well assume that the discount function was a function of time, or even a stochastic process. For simplicity, we will not pursue this, but see Norberg (1991) and Møller (1998).

4.7 *Bases and Expected Present Values*

In terms of probability models, all we have defined so far are the elements of the sample space Ω (the sample paths $N_{01}(t)$) and some related functions such as L and $V(s)$. We have not introduced any σ-algebras, filtrations or probability measures, nor have we carried out any probabilistic calculation, such as taking expectations. We now consider these:

(a) Our filtration is the 'natural' filtration generated by the process $N_{01}(t)$, which is easily described. At time t, the past values $N_{01}(s)$ $(s \leq t)$ are all known, and the future values $N_{01}(s)$ $(s > t)$ are unknown (unless $N_{01}(t) = 1$, in which case nothing more can happen). This information is summed up by the σ-algebra $\mathcal{F}_t$.
To picture this filtration, cover Figure 2 with your hand, and then slowly reveal the life history. Before age 46, all possible future life histories are hidden by your hand; the information $\mathcal{F}_t$ is the combination of the revealed life history and all these hidden possibilities.

(b) Our 'overall' σ-algebra $\mathcal{F}$ is the union of all the $\mathcal{F}_t$.

(c) The **probability measure** corresponds to the **mortality basis**. As is well known, the actuary will choose a different mortality basis for different purposes, and we suppose that nature chooses the 'real' mortality basis. In other words, the sample space and the filtration do not determine the choice of probability measure; nor is the choice of probability measure always an attempt to find nature's 'real' probabilities (that is the estimation problem). This point is of even greater importance in financial mathematics, where it is often misunderstood.

All concrete calculations depend on the choice of probability measure (mortality basis). We will illustrate this using expected present values. Suppose the actuary has chosen a probability measure P (equivalent to life table probabilities ${}_tp_x$). Taking as an example the whole life assurance benefit, for a life aged x, say, $\mathrm{E}_P[X]$ is:

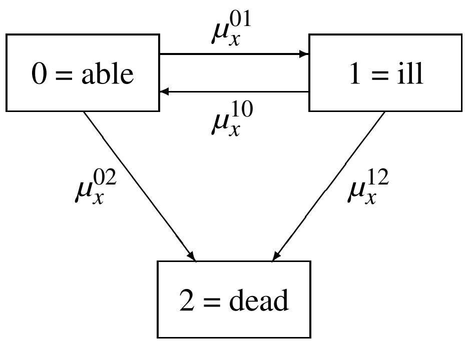

Figure 3: An illness-death model

$$\mathrm{E}_P\left[\int_0^\infty v^t dN_{01}(t)\right] = \int_0^\infty v^t \mathrm{E}_P[dN_{01}(t)] = \int_0^\infty v^t \mathrm{P}[dN_{01}(t) = 1] = \int_0^\infty v^t\, {}_tp_x \mu_{x+t} dt \qquad (10)$$

which should be familiar[5]. If the actuary chooses a different measure P^*, say (equivalent to different life table probabilities ${}_tp_x^*$), we get a different expected value:

$$\mathrm{E}_{P^*}[X] = \int_0^\infty v^t\, {}_tp_x^* \mu_{x+t}^* dt. \qquad (11)$$

Expected values of annuities are also easily written:

$$\mathrm{E}_P[Y] = \int_0^\infty v^t\, {}_tp_x dt. \qquad (12)$$

4.8 *More Examples of Counting Processes*

Figure 3 shows the well-known illness-death model. A precise formulation begins with the state $S(t)$ occupied at time t; a stochastic process.

Figure 4 shows a single sample path from $S(t)$: a life who has a short illness at age 40, recovers at age 42, then has a longer, ultimately fatal illness starting at age 49. In the 2-state mortality model, the stochastic process $S(t)$, representing the state occupied, coincided with the counting process $N_{01}(t)$ representing the number of events[6]: here it is not so. In fact we can define 4 counting processes, one for each transition, for example:

$$N_{01}(t) \quad = \quad \text{No. of transitions able to ill}$$

[5]The last step in equation (10) follows because the event $\{dN_{01}(t) = 1\}$ is just the event 'survives to *just before* age $x+t$, then dies in the next instant', which has the probability ${}_tp_x\mu_{x+t}dt$.

[6]We did not introduce $S(t)$ for the 2-state model: we do so now, it is the same as $N_{01}(t)$.

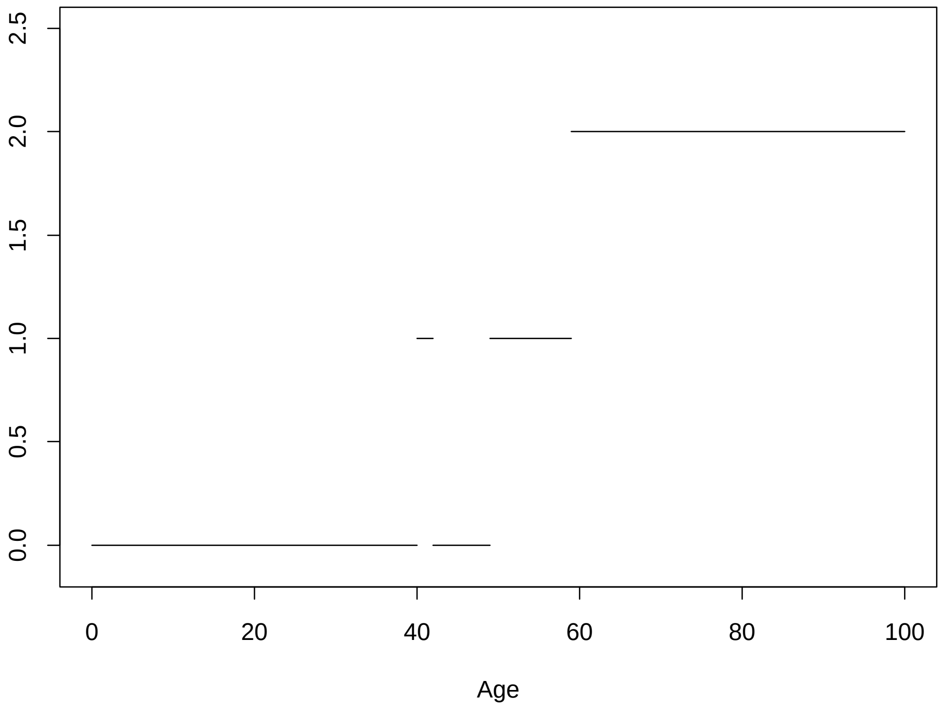

Figure 4: A sample path of an illness-death process $S(t)$: 0=able, 1=ill, 2=dead

$$
\begin{aligned}
N_{02}(t) &= \text{No. of transitions able to dead} \\
N_{10}(t) &= \text{No. of transitions ill to able} \\
N_{12}(t) &= \text{No. of transitions ill to dead}
\end{aligned}
$$

or, regarding them as one object, we have a multivariate counting process with 4 components. We can also define stochastic processes indicating presence in each state, $I_j(t)$, annuity payment functions $a_j(t)$ for each state, and sum assured functions for each possible transition, $A_{jk}(t)$. Then all of the life insurance mathematics from the 2-state model carries over with only notational changes.

4.9 *Where are the Martingales?*

We have not yet mentioned any martingales associated with counting processes, but they are very simple, and central to both data analysis and applications. In the 2-state model, the martingale is:

$$M_{01}(t) = N_{01}(t) - \int_0^t I_0(s)\mu_s ds. \tag{13}$$

$M_{01}(t)$ is called the **compensated** counting process, and the integral on the right hand side is called the **compensator** of $N_{01}(t)$. It is easy to see that $M_{01}(t)$ is a martingale from its increments:

$$dM_{01}(t) \quad = \quad dN_{01}(t) - I_0(t)\mu_t dt \tag{14}$$

$$\mathrm{E}_P[dM_{01}(t)] \;=\; \mathrm{E}_P[dN_{01}(t)] - \mathrm{E}_P[I_0(t)\mu_t dt] = 0 \tag{15}$$

We have been careful to specify the probability measure P in the expectation. If we change the measure, for example to P^*, corresponding to probabilities ${}_tp^*_x$, we get a different martingale:

$$M^*_{01}(t) = N_{01}(t) - \int_0^t I_0(s)\mu^*_s ds \tag{16}$$

and $\mathrm{E}_{P^*}[dM^*_{01}(t)] = 0$. Alternatively, given a force of mortality μ^*_t, we can find a probability measure P^* such that $M^*_{01}(t)$ is a P^*-martingale; P^* is simply given by the probabilities ${}_tp^*_x = \exp(-\int_0^t \mu^*_s ds)$. This is true of any (well-behaved) force of mortality, not just nature's chosen 'true' force of mortality[7].

An idea of the usefulness of $M_{01}(t)$ can be gained from equation (13). If we consider an age interval short enough that a constant transition intensity μ is a reasonable approximation, this becomes:

$$M_{01}(t) = N_{01}(t) - \mu \int_0^t I_0(s)ds. \tag{17}$$

But the two random quantities on the right are just the number of deaths $N_{01}(t)$, and the total time spent at risk $\int_0^t I_0(s)ds$, better known as the central exposed to risk. All the properties of the maximum likelihood estimate of μ, based on these two statistics (summed over many independent lives) are consequences of the fact that $M_{01}(t)$ is a martingale (see Macdonald (1996a, 1996b)).

For more complicated models, we get a set of martingales, one for each possible transition (from state j to state k) of the form:

$$M_{jk}(t) = N_{jk}(t) - \int_0^t I_j(s)\mu^{jk}_s ds \tag{18}$$

which have all the same properties.

4.10 *Prospective and Retrospective Reserves*

We now return to equation (9): $V(s) = v^{-s}\int_0^T v^t dL(t)$. Recall that the premium is part of the payment function $a_0(t)$; setting the premium according to the **equivalence principle** simply means setting $\mathrm{E}_P[V(0)] = 0$ and solving for $a_0(t)$, where P is the probability measure corresponding to the premium basis.

For convenience, we will use the same basis (measure) for premiums and reserves, as is common in other European countries.

Reserves follow when we consider the evolution of the value function V over time, as information emerges. We start from the conditional expectation; for $s < T$:

[7] This is exactly analogous to the 'equivalent martingale measure' of financial mathematics, in which we are given the drift of a geometric Brownian motion (coincidentally, also often denoted μ_t) and then find a probability measure under which the discounted process is a martingale.

$$\begin{aligned} \mathrm{E}_P[V(s)|\mathcal{F}_s] &= \mathrm{E}_P\left[\frac{1}{v^s}\int_0^T v^t\,dL(t)\,\middle|\,\mathcal{F}_s\right] &(19)\\ &= \mathrm{E}_P\left[\frac{1}{v^s}\int_0^s v^t\,dL(t)\,\middle|\,\mathcal{F}_s\right]+\mathrm{E}_P\left[\frac{1}{v^s}\int_s^T v^t\,dL(t)\,\middle|\,\mathcal{F}_s\right] &(20)\end{aligned}$$

The second term on the right is the **prospective reserve** at time s. If the information $\mathcal{F}_s$ is the complete life history up to time s, it is the same as the usual prospective reserve. However, this definition is more general; for example, under a joint-life second-death assurance, the first death might not be reported, so that $\mathcal{F}_s$ represents incomplete information. Also, it does not depend on the probabilistic nature of the process generating the life history; it is not necessary to suppose that the process is Markov, for example. If the process is Markov (as we often suppose) then conditioning on $\mathcal{F}_s$ simply means conditioning on the state occupied at time s, which is very convenient in practice.

The first term on the right is minus the **retrospective reserve**. This definition of the retrospective reserve is new (Norberg, 1991) and is not equivalent to 'classical' definitions. This is a striking achievement of the stochastic process approach: for convenience we also list some of the notions of retrospective reserve that have preceded it:

(a) The 'classical' retrospective reserve (for example, Neill (1977)) depends on a deterministic cohort of lives, who share out a fund among survivors at the end of the term. However, this just exposes the weaknesses of the deterministic model: given a whole number of lives at outset, l_x say, the number of survivors some time later, $l_{x\,t}p_x$ is usually not an integer. Viewed prospectively this can be excused as being a convenient way of thinking about expected values, but viewed retrospectively there is no such excuse.

(b) Hoem (1969) allowed both the number of survivors, and the fund shared among survivors, to be random, and showed that the classical retrospective reserve was obtained in the limit, as the number of lives increased to infinity.

(c) Perhaps surprisingly, the 'classical' notion of retrospective reserve does not lead to a unique specification of what the reserve should be in each state of a general Markov model, leading to several alternative definitions (Hoem, 1988; Wolthius & Hoem, 1990; Wolthius, 1992) in which the retrospective and prospective reserves in the initial state were equated by definition.

(d) Finally, Norberg (1991) pointed out that the 'classical' retrospective reserve is "... rather a retrospective formula for the prospective reserve ...", and introduced the definition in equation (20). This is properly defined for individual lives, and depends on known information $\mathcal{F}_s$. If $\mathcal{F}_s$ is the complete life history, the conditional expectation disappears and:

$$\text{Retrospective reserve} = \frac{-1}{v^s}\int_0^s v^t\,dL(t) \tag{21}$$

which is more akin to an asset share on an individual policy basis. If $\mathcal{F}_s$ represents coarser information, for example aggregate data in respect of a cohort of policies, the retrospective reserve is akin to an asset share with pooling of mortality costs.

We have spent some time on retrospective reserves, because it is an example of the greater clarity obtained from a careful mathematical formulation of the process being modelled, in this case the life history.

4.11 *Differential Equations*

The chief computational tools associated with multiple-state models are **ordinary differential equations** (ODEs). We mention three useful systems of ODEs:

(a) The **Kolmogorov forward equations** can be found in any textbook on Markov processes (for example, Kulkarni (1995)) and have been in the actuarial syllabus for some time. They allow us to calculate transition probabilities in a Markov process, given the transition intensities, which is exactly what we need since transition intensities are the quantities most easily estimated from data. We give just one example, the simplest of all from the 2-state model:

$$\frac{\partial}{\partial t}{}_t p_x = -{}_t p_x \mu_{x+t}. \tag{22}$$

(b) **Theile's equation** governs the development of the prospective reserve. For example, if ${}_t\overline{V}_x$ is the reserve under a whole life assurance for £1, Theile's equation is:

$$\frac{d}{dt}{}_t\overline{V}_x = \delta_t \overline{V}_x + \overline{P}_x - (1 - {}_t\overline{V}_x)\mu_{x+t} \tag{23}$$

which has a very intuitive interpretation. In fact, it is the continuous-time equivalent of the recursive formula for reserves well-known to British actuaries. It was extended to any Markov model by Hoem (1969).

(c) Norberg (1995b) extended Theile's equations for prospective policy values (that is, first moments of present values) to second and higher moments. We do not show these equations, as that would need too much new notation, but we note that they were obtained from the properties of counting process martingales.

Most systems of ODEs do not admit closed-form solutions, and have to be solved numerically, but many methods of solution are quite simple[8], and well within the capability of a modern PC. So, while closed-form solutions are nice, they are not too important, and it is better to seek ODEs that are relevant to the problem, rather than explicitly soluble. We would remind actuaries of a venerable example of a numerical solution to an intractable ODE, namely the life table.

4.12 *Advantages of the Counting Process Approach*

(a) First and foremost, counting processes represent complete life histories. In practice, not all this information might be available or useable, but it is best to start with a model that represents the underlying process, and then to make whatever approximations might be needed to meet the circumstances (for example, data grouped into years).

(b) The mathematics of counting processes and multiple-state models is easily introduced in terms of the 2-state mortality model, but carries over to any more complicated model, thus solving problems that defeat life-table methods. This is increasingly important in practice, as new insurances are introduced.

[8] Numerical solution of ODEs is one of the most basic tasks in numerical analysis.

(c) Completely new results have been obtained, such as an operational definition of retrospective reserves, and Norberg's differential equations.
(d) The tools we use are exactly those that are essential in modern financial mathematics, in particular stochastic integrals and conditional expectations. For a remarkable synthesis of these two fields, see Møller (1998). An alternative approach, in which rates of return as well as lifetimes are modelled by Markov processes, has been developed (Norberg, 1995b) extending greatly the scope of the material discussed here.
(e) We have not discussed data analysis, but mortality studies are increasingly turning towards counting process tools, for exactly the same reason as in (a). It will often be helpful for actuaries at least to understand the language.

5. Finance

5.1 *Introduction*

In this section we are going to illustrate how stochastic processes can be used to price **financial derivatives**.

A financial derivative is a contract which derives its value from some underlying security. For example, a European call option on a share gives the holder the right, but not the obligation, to buy the share at the exercise date T at the strike price of K. If the share price at time T, S_T, is less than K then the option will not be exercised and it will expire without any value. If S_T is greater than K then the holder will exercise the option and a profit of $S_T - K$ will be made. The profit at T is, therefore, $\max\{S_T - K, 0\}$.

5.2 *Models of Asset Prices*

Much of financial mathematics must be based on explicit models of asset prices, and the results we get depend on the models we decide to use. In this section we will look at two models for share prices: a simple binomial model which will bring out the main points; and geometric Brownian motion. Throughout we make the following general assumptions[9].

(a) We will use S_t to represent the price of a non-dividend-paying stock at time t ($t = 0, 1, 2, \ldots$). For $t > 0$, S_t is random.
(b) Besides the stock we can also invest in a bond or a cash account which has value B_t at time t per unit invested at time 0. This account is assumed to be risk free and we will assume that it earns interest at the constant risk-free continuously compounding rate of r per annum. Thus $B_t = \exp(rt)$. (In discrete time, **risk free** means that we know at time $t-1$ what the value of the risk-free investment will be at time t. In this more simple case, the value of the risk-free investment at any time t is known at time 0.)
(c) At any point in time we can hold arbitrarily large amounts (positive or negative) of stock or cash.

5.3 *The No-Arbitrage Principle*

Before we progress it is necessary to discuss **arbitrage**.

[9] These assumptions can be relaxed considerably with more work.

Suppose that we have a set of assets in which we can invest (with holdings which can be positive or negative). Consider a particular portfolio which starts off with value zero at time 0 (so we have some positive holdings and some negative). With this portfolio, it is known that there is some time T in the future when its value will be non-negative with certainty and strictly positive with probability greater than zero. This is called an arbitrage opportunity. To exploit it we could multiply up all amounts by one thousand or one million and make huge profits without any cost or risk.

In financial mathematics and derivative pricing we make the fundamental assumption that arbitrage opportunities like this do not exist (or at least that if they do exist, they disappear too quickly to be exploited).

5.4 *A One-Period Binomial Model*

First we consider a model for stock prices over **one discrete time period**. We have two possibilities for the price at time 1 (see Figure 5):

$$S_1 = \begin{cases} S_0 u & \text{if the price goes up} \\ S_0 d & \text{if the price goes down} \end{cases}$$

with $d < u$ (strictly, it is not necessary that $d < 1$).

In order to avoid arbitrage we must have $d < e^r < u$. Suppose this is not the case: for example, if $e^r < d$. Then we could borrow £1 of cash and buy £1 of stock. At time 0 this would have a net cost of £0. At time 1 our portfolio would be worth $d - e^r$ or $u - e^r$ both of which are greater than 0: an example of arbitrage.

Suppose that we have a derivative which pays f_u if the price of the **underlying** stock goes up and f_d if the price of the underlying stock goes down. At what price should this derivative trade at time 0?

In this model (and also in the multi-period model that we consider later) we will assume:

(a) there are no trading costs;
(b) there are no minimum or maximum units of trading;
(c) stock and bonds can only be bought and sold at discrete times 1, 2, ...

As such the model appears to be quite unrealistic. However, it does provide us with good insight into the theory behind more realistic models. Furthermore it provides us with an effective computational tool for derivatives pricing.

At time 0 suppose we hold ϕ units of stock and ψ units of cash. The value of this portfolio at time 0 is V_0. At time 1 the same portfolio has the value:

$$V_1 = \begin{cases} \phi S_0 u + \psi e^r & \text{if the stock price goes up} \\ \phi S_0 d + \psi e^r & \text{if the stock price goes down} \end{cases}$$

Let us choose ϕ and ψ so that $V_1 = f_u$ if the stock price goes up and $V_1 = f_d$ if the stock price goes down. Then:

$$\begin{aligned} \phi S_0 u + \psi e^r &= f_u \\ \text{and} \quad \phi S_0 d + \psi e^r &= f_d \end{aligned}$$

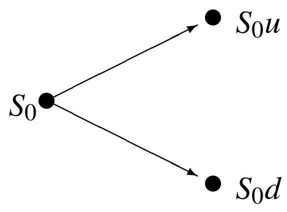

Figure 5: One-period binomial model for stock prices

Thus we have two linear equations in two unknowns, ϕ and ψ. We solve this system of equations and find that:

$$\begin{aligned}
\phi &= \frac{f_u - f_d}{S_0(u-d)} \\
\psi &= e^{-r}(f_u - \phi S_0 u) \\
&= e^{-r}\left(f_u - \frac{(f_u - f_d)u}{u-d}\right) \\
&= e^{-r}\left(\frac{f_d u - f_u d}{u-d}\right) \\
\Rightarrow V_0 &= \phi S_0 + \psi \\
&= \frac{(f_u - f_d)}{u-d} + e^{-r}\frac{(f_d u - f_u d)}{u-d} \\
&= f_u\left(\frac{1 - de^{-r}}{u-d}\right) + f_d\left(\frac{-1 + ue^{-r}}{u-d}\right) \\
&= e^{-r}(qf_u + (1-q)f_d) \\
\text{where } q &= \frac{e^r - d}{u-d} \\
1-q &= \frac{u-e^r}{u-d} = 1 - \frac{e^r - d}{u-d}
\end{aligned}$$

Note that the no-arbitrage condition $d < e^r < u$ ensures that $0 < q < 1$.

If we denote the payoff of the derivative at $t = 1$ by the random variable $f(S_1)$, we can write:

$$V_0 = e^{-r}\mathrm{E}_Q(f(S_1))$$

where Q is a probability measure which gives probability q to an upward move in prices and $1-q$ to a downward move. We can see that q depends only upon u, d and r and not upon the potential derivative prices. In particular, Q does not depend on the type of derivative; it is the same for all derivatives on the same stock.

The portfolio (ϕ, ψ) is called a **replicating portfolio** because it replicates, precisely, the payoff at time 1 on the derivative without any risk. It is also a simple example of a hedging strategy: that is, an investment strategy which reduces the amount of risk carried by the issuer of the contract. In this respect not all hedging strategies are replicating strategies.

Up until now we have not mentioned the **real-world** probabilities of up and down moves in prices. Let these be p and $1-p$ where $0 < p < 1$, defining a probability measure P.

Other than by total coincidence, p will not be equal to q.

Let us consider the expected stock price at time 1. Under P this is:

$$S_0(pu + (1-p)d) = \mathrm{E}_P(S_1)$$

and under Q it is:

$$\mathrm{E}_Q(S_1) = S_0(qu + (1-q)d) = S_0\left(\frac{u(e^r - d)}{u-d} + \frac{d(u-e^r)}{u-d}\right) = S_0 e^r.$$

Under Q we see that the expected return on the **risky** stock is the same as that on a risk-free investment in cash. In other words under the probability measure Q investors are neutral with regard to risk: they require no additional returns for taking on more risk. This is why Q is sometimes referred to as a **risk-neutral probability measure**.

Under the real-world measure P the expected return on the stock will not normally be equal to the return on risk-free cash. Under normal circumstances investors demand higher expected returns in return for accepting the risk in the stock price. Thus we would normally find that $p > q$. However, this makes no difference to our analysis.

5.5 *Comparison of Actuarial and Financial Economic Approaches*

The actuarial approach to the pricing of this contract would give:

$$V_0^a = e^{-\delta}\mathrm{E}_P[f(S_1)] = e^{-\delta}(pf_u + (1-p)f_d)$$

where δ is the actuarial, risk-discount rate. Compare this with the price calculated using the principles of financial economics above:

$$V_0 = e^{-r}\mathrm{E}_Q(f(S_1)) = e^{-r}(qf_u + (1-q)f_d).$$

If forwards are trading at V_0^a, where $V_0^a > V_0$, then we can sell one derivative at the actuarial price, and use an amount V_0 to set up the replicating portfolio (ϕ, ψ) at time 0. The replicating portfolio ensures that we have the right amount of money at $t = 1$ to pay off the holder of the derivative contract. The difference between V_0^a and V_0 is then guaranteed profit with no risk.

Similarly if $V_0^a < V_0$ we can also make arbitrage profits.

(In fact neither of these situations could persist for any length of time because demand for such contracts trading at V_0^a would push the price back towards V_0 very quickly. This is a fundamental principle of financial economics: that is, prices should not admit arbitrage opportunities. If they did exist then the market would spot any opportunities very quickly and the resulting excess supply or demand would remove the arbitrage opportunity before any substantial profits could be made. In other words, arbitrage opportunities might exist for very short periods of time in practice, while the market is free from arbitrage for the great majority of time and certainly at any points in time where large financial transactions are concerned. Of course, we would have no problem in buying such a contract if we were to offer a price of V_0^a to the seller if this was greater than V_0 but we would not be able to sell at that price. Similarly we could easily sell such a contract if $V_0^a < V_0$ but not buy at that price. In both cases we would be left in a position where we would have to maintain a risky portfolio in order to give ourselves a chance of a profit, since hedging would result in a guaranteed loss.)

For V_0^a to make reasonable sense, then, we must set δ in such a way that V_0^a equals V_0. In other words, the subjective choice of δ in actuarial work equates to the objective selection of the risk-neutral probability measure Q. Choosing δ to equate V_0^a and V_0 is not what happens in practice and, although δ is set with regard to the level of risk under the derivative contract, the subjective element in this choice means that there is no guarantee that V_0^a will equal V_0. In general, therefore, the actuarial approach, on its own, is not appropriate for use in derivative pricing. Where models are generalised and assumptions weakened to such an extent that it is not possible to construct hedging strategies which **replicate** derivative payoffs then there is a role for a combination of the financial economic and actuarial approaches. However, this is beyond the scope of this paper.

5.6 *Binomial Lattices*

Now let us look at how we might price a derivative contract in a multiperiod model with n time periods. Let $f(x)$ be the payoff on the derivative if the share has a price of x at the expiry date n. For example, for a European call option we have $f(x) = \max\{x-K, 0\}$, where K is the strike price.

Suppose now that over each time period the share price can rise by a factor of u or fall by a factor of $d = 1/u$: that is, for all t, S_{t+1} is equal to $S_t u$ or $S_t d$. This means that the effect of successive 'up and down' moves is the same as successive 'down and up' moves. Furthermore the risk-free rate of interest is constant and equal to r, with, still, $d < e^r < u$. Then we have:

$$S_t = S_0 u^{N_t} d^{t-N_t}$$

where N_t is the number of up-steps[10] between time 0 and time t. This means that we have $n+1$ possible states at time n. We can see that the value of the stock price at time t depends only upon the number of up and down steps and not on the order in which they occurred. Because of this property the model is called a **recombining binomial tree** or a **binomial lattice** (see Figure 6).

The sample space for this model, Ω, is the set of all sample paths from time 0 to time n. This is widely known as the random walk model.There are 2^n such sample paths since there are two pssible outcomes in each time period. The information $\mathcal{F}$ is the σ-algebra generated by all sample paths from time 0 to n while the filtrations $\mathcal{F}_t$ are generated by all sample paths up to time t. (Given the sample space Ω, each sample path up to time t is equivalent to 2^{n-t} elements of the sample space, each element being the same over the period 0 to t. N_t and S_t are random variables which are functions of the sample space.)

Under this model all periods have the same probability of an up step and steps in each time period are independent of one another. Thus the number of up steps up to time t, N_t, has under

[10]In this sense, N_t can also be regarded as a discrete-time counting process; see Section 4.

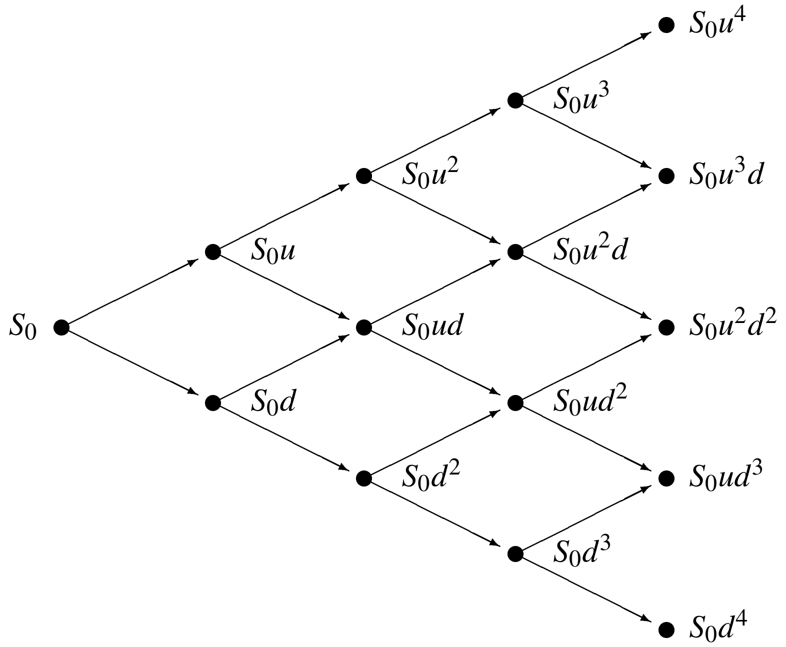

Figure 6: Recombining binomial tree or binomial lattice

Q a binomial distribution with parameters t and q. Furthermore, for $0 < t < n$, N_t is independent of $N_n - N_t$ and $N_n - N_t$ has a binomial distribution with parameters $n - t$ and q.

Let us extend our notation a little bit. Let $V_t(j)$ be the fair value of the derivative at time t given $N_t = j$ for $j = 0, \ldots, t$. Also let $V_n(j) = f\left(S_0 u^j d^{n-j}\right)$. Finally we write $V_t = V_t(N_t)$ to be the random value at some future time t.

In order for us to calculate the value at time 0, $V_0(0)$, we must work backwards one period at a time from time n making use of the one-period binomial model as we go.

First let us consider the time period $n-1$ to n. Suppose that $N_{n-1} = j$. Then, by analogy with the one-period model we have:

$$
\begin{aligned}
V_{n-1}(j) &= e^{-r}\left[qV_n(j+1) + (1-q)V_n(j)\right] \\
&= e^{-r}\mathrm{E}_Q\left[V_n \mid \mathcal{F}_{n-1}\right] \\
&= e^{-r}\mathrm{E}_Q\left[f(S_n) \mid N_{n-1} = j\right] \\
&= e^{-r}\mathrm{E}_Q\left[f(S_n) \mid \mathcal{F}_{n-1}\right] \\
\text{where } q &= \frac{e^r - d}{u - d}.
\end{aligned}
$$

Equivalently we can write this as $V_{n-1} = e^{-r}\mathrm{E}_Q[f(S_n) \mid \mathcal{F}_{n-1}]$.

As we work backwards we have:

$$\begin{aligned}
V_{t-1} &= e^{-r}\mathrm{E}_Q[V_t \mid \mathcal{F}_{t-1}] \\
&= e^{-r}\mathrm{E}_Q\left[e^{-r}\mathrm{E}_Q(V_{t+1} \mid \mathcal{F}_t) \mid \mathcal{F}_{t-1}\right] \\
&= e^{-2r}\mathrm{E}_Q[V_{t+1} \mid \mathcal{F}_{t-1}] \qquad \text{(using the Tower Law)} \\
&\vdots \qquad \vdots \qquad \vdots \\
&= e^{-(n-t+1)r}\mathrm{E}_Q[V_n \mid \mathcal{F}_{t-1}] \\
&= e^{-(n-t+1)r}\mathrm{E}_Q[f(S_n) \mid \mathcal{F}_{t-1}].
\end{aligned}$$

Finally we get to:

$$V_0 = e^{-nr}\mathrm{E}_Q[f(S_n) \mid \mathcal{F}_0] = e^{-nr}\mathrm{E}_Q[f(S_n) \mid S_0].$$

The price at time t of the path-dependent derivative is thus:

$$\begin{aligned}
V_t &= e^{-r(n-t)}\mathrm{E}_Q\left[f\left(S_t u^{N_n-N_t} d^{(n-t)-(N_n-N_t)}\right) \mid N_t\right] \\
&= e^{-r(n-t)}\sum_{k=0}^{n-t} f\left(S_t u^k d^{n-t-k}\right)\frac{(n-t)!}{k!(n-t-k)!}q^k(1-q)^{n-t-k}.
\end{aligned}$$

We have noted before that $\mathrm{E}_Q(S_1) = S_0 e^r$ giving rise to the use of the name **risk-neutral measure** for Q. Similarly in the n-period model we have (putting $f(s) = s$):

$$\mathrm{E}_Q[S_t|\mathcal{F}_0] = S_0 e^{rt}.$$

So the use of the expression **risk-neutral measure** for Q is still valid. Alternatively we can write:

$$\mathrm{E}_Q\left[e^{-rT}S_T \mid \mathcal{F}_t\right] = e^{-rt}S_t.$$

In other words, the discounted asset value process $D_t = e^{-rt}S_t$ is a martingale under Q. This gives rise to another name for Q: **equivalent martingale measure**.

In fact we normally use this result the other way round, as we will see in the next section. That is, the first thing we do is to find the equivalent martingale measure Q, and then use it immediately to price derivatives.

5.7 *A Continuous-Time Model*

Let us now work in continuous time. Let S_t be the price of the non-dividend-paying share for $0 \le t \le T$. Suppose that a derivative pays $f(s)$ at time T if the share price at time T is equal to s.

The particular model we are going to look at for S_t is called **geometric Brownian motion**: that is, $S_t = S_0 \exp[(\mu - \frac{1}{2}\sigma^2)t + \sigma Z_t]$ where Z_t is a standard Brownian motion under the real-world measure P. (For the properties of Brownian motion see Appendix A.) This means that S_t has a log-normal distribution with mean $S_0 \exp(\mu t)$ and variance $\exp(2\mu t).\left[\exp(\sigma^2 t) - 1\right]$. By application of Itô's lemma (see Appendix B) we can write down the **stochastic differential equation** (SDE) for S_t as follows:

$$dS_t = \mu S_t dt + \sigma S_t dZ_t.$$

By analogy with the binomial model there is another probability measure Q (the risk-neutral measure or equivalent martingale measure) under which:

(a) $e^{-rt}S_t$ is a martingale

(b) S_t can be written as the geometric Brownian motion $S_0 \exp\left[\left(r - \frac{1}{2}\sigma^2\right)t + \sigma\tilde{Z}_t\right]$ where $\tilde{Z}(t)$ is a standard Brownian motion under Q

By continuing the analogy with the binomial model (for example, see Baxter & Rennie (1996)) we can also say that the value at time t of the derivative is:

$$V_t = e^{-r(T-t)}\mathrm{E}_Q\left[f(S_T) \mid \mathcal{F}_t\right] = e^{-r(T-t)}\mathrm{E}_Q\left[f(S_T) \mid S_t\right].$$

With a bit more work we can also see that, under this model, if we invest V_t in the right way (that is, with a suitable hedging strategy), then we can replicate the payoff at T without the need for extra cash.

Suppose that we consider a European call option, so that $f(s) = \max\{s - K, 0\}$. Then we can exploit a well known property of the log-normal distribution to get the celebrated Black-Scholes formula:

$$\begin{aligned} V_t &= S_t\Phi(d_1) - Ke^{-r(T-t)}\Phi(d2) \\ \text{where } d_1 &= \frac{\log\frac{S_t}{K} + \left(r + \frac{1}{2}\sigma^2\right)(T-t)}{\sigma\sqrt{T-t}} \\ \text{and } d_2 &= d_1 - \sigma\sqrt{T-t}. \end{aligned}$$

A more detailed development of pricing and hedging of derivatives in continuous time can be found in Baxter & Rennie (1996).

6. Applications in Risk Theory

6.1 *Introduction*

In this section we show how stochastic processes can be used to gain insight into the pricing of general insurance policies. In particular, we will make use of the notion of a martingale, Brownian motion and also the optional stopping theorem.

Suppose we have a general insurance risk, for example comprehensive insurance for a fleet of cars or professional indemnity insurance for a software supplier, for which cover is required on an annual basis. We want to answer the question: "*In an ideal world, how should* *we calculate the annual premium for this cover?*" Let us denote by P_n the premium to be charged for cover in year n, where the coming year is year 1.

The starting point is to consider the claims which will arise each year. Since the aggregate amount of claims is uncertain, we model these amounts as random variables. Let S_n be a random variable denoting the aggregate claims arising from the risk in year n. We might also take into consideration, particularly for a large risk, the amount of capital with which we are prepared to back this risk. We denote this initial capital, or 'initial surplus', U. In practice, we would also take into consideration many other factors, for example, expenses and the premium rates charged by our competitors. However, to keep things simple, we will ignore these other factors.

Throughout this section we will assume that the random variables $\{S_n\}_{n=1}^{\infty}$ are independent of each other, but we will not assume they are identically distributed. To be able to calculate P_n we need to be able to calculate, or at least estimate, the distribution of S_n. If we have information about the distributions of claim numbers and claim amounts in year n, we may be able to use Panjer's celebrated recursion formula to calculate the distribution of S_n. See, for example, Klugman *et al* (1997). In some circumstances, for example, when S_n is the sum of a large number of independent claim amounts, it may be reasonable to assume that S_n has, approximately, a normal distribution. In what follows we will occasionally make the following assumption:

$$S_n \sim N(\mu_n, \sigma_n^2) \tag{24}$$

for some parameters μ_n and σ_n.

6.2 *The Standard Deviation Principle*

Suppose we decide that each year P_n should be set at a level such that the probability that aggregate claims exceed the premium in that year should be suitably small, say $1-p$. Formally, this criterion can be expressed as follows:

$$\mathrm{P}[S_n < P_n] = p. \tag{25}$$

Making the additional assumption that S_n is normally distributed, that is, assumption (24), it is easy to see that:

$$P_n = \mu_n + \gamma_p \sigma_n \tag{26}$$

where γ_p is such that:

$$\Phi(\gamma_p) = p$$

and $\Phi(z)$ is the cumulative distribution function of the $N(0,1)$ distribution.

Formula (26) says that in year n, P_n should be calculated as the mean of the aggregate claims for that year plus a loading proportional to the standard deviation of the aggregate claims. Notice that the proportionality factor, γ_p, does not depend on n. In the actuarial literature, formula (26) is known as the *standard deviation principle* for the calculation of premiums, and γ_p is referred to as the loading factor. See Goovaerts, De Vylder, & Haezendonck (1984).

6.3 *Utility Functions and the Variance Principle*

Now suppose that our attitude to money is summarised by a utility function, $u(x)$. Intuitively, $u(x)$ is a real-valued function which expresses 'how much we like an amount of money x'. Mathematically, we require the following two conditions to hold:

$$\frac{d}{dx}u(x) > 0 \quad \text{and} \quad \frac{d^2}{dx^2}u(x) < 0.$$

The first of these conditions says that "we always prefer more money". The second condition says that "an extra pound is worth less, the wealthier we are". See Bowers *et al.* (1997) for more details.

We can use this utility function to calculate P_n from the following formula:

$$u(W) = \mathrm{E}[u(W + P_n - S_n)] \tag{27}$$

where W is our wealth at the start of the n-th year. The rationale behind this formula is as follows: if we do not insure this risk, the utility of our wealth is $u(W)$; if we do insure the risk, our expected utility of wealth at the end of the year is $\mathrm{E}[u(W + P_n - S_n)]$. Formula (27) says that for a premium of P_n we are indifferent between insuring the risk and not insuring it. In this sense the premium P_n is 'fair'. See, for example, Bowers *et al.* (1997) for details.

Now assume that $u(x)$ is an **exponential utility function** with parameter $\alpha > 0$, so that:

$$u(x) = 1 - e^{-\alpha x}. \tag{28}$$

Using (28) in (27) and solving for P_n, we have:

$$P_n = \alpha^{-1} \log\left(M_{S_n}(\alpha)\right) \tag{29}$$

where the notation $M_Z(.)$ denotes the moment generating function of a random variable Z, so that $M_Z(\alpha) = \mathrm{E}[e^{\alpha Z}]$. Notice that in this special case P_n does not depend on W, our wealth at the start of the n-th year.

Finally, let us assume that S_n has a normal distribution, as in (24). Then:

$$M_{S_n}(\alpha) = \exp\{\mu_n \alpha + \frac{1}{2}\alpha^2 \sigma_n^2\}$$

and so (29) becomes:

$$P_n = \mu_n + \frac{1}{2}\alpha\sigma_n^2. \tag{30}$$

Thus, P_n is calculated according to the *variance principle*.

6.4 *Multi-period Analysis — Discrete Time*

Formulae (26) and (30) provide two alternative ways of calculating P_n in an ideal world. They have some features in common:

(a) in each case P_n is the sum of the expected value of S_n and a positive loading; and
(b) in each case we arrived at a formula for P_n by considering the n-th year in isolation from any other years.

A major difference in the development so far is that for (26) the loading factor γ_p has an intuitive meaning, whereas in (30), or (29), the parameter α is not so easily understood or quantified. To fill in this gap, we need to consider our surplus at the end of each year in the future.

Let U_n denote the surplus at the end of the n-th year, so that:

$$U_n = U + \sum_{k=1}^{n} (P_k - S_k)$$

for $n = 1,2,3,\ldots$, with U_0 defined to be U. Now define:

$$\psi(U) = \mathrm{P}[U_n \le 0 \quad \text{for some } n,\ n = 1,2,3,\ldots]$$

to be the probability that at some time in the future we will need more capital to back this risk, that is, in more emotive language, the probability of ultimate ruin in discrete time for our surplus process. Let us assume that P_n is calculated using (29).

Consider the process $\{Y_n\}_{n=0}^{\infty}$, where $Y_n = \exp\{-\alpha U_n\}$ for $n = 0,1,2,\ldots$. Then $\{Y_n\}_{n=0}^{\infty}$ is a martingale with respect to $\{S_n\}_{n=1}^{\infty}$. To prove this, note that:

$$U_{n+1} = U_n + P_{n+1} - S_{n+1}$$

implies that:

$$\begin{aligned} & E\left[Y_{n+1}|S_1,\ldots,S_n\right] \\ = \;& E\left[\exp\{-\alpha\left(U_n + P_{n+1} - S_{n+1}\right)|S_1,\ldots,S_n\}\right] \\ = \;& E\left[\exp\{-\alpha(P_{n+1} - S_{n+1})\}|S_1,\ldots,S_n\right]\exp\{-\alpha U_n\}. \end{aligned}$$

Since $\{S_n\}_{n=1}^{\infty}$ is a sequence of independent random variables it follows that:

$$\begin{aligned} & E\left[\exp\{-\alpha(P_{n+1} - S_{n+1})\}|S_1,\ldots,S_n\right] \\ = \;& \mathrm{E}[\exp\{-\alpha(P_{n+1} - S_{n+1})\}] \\ = \;& \exp\{-\alpha P_{n+1}\}E\left[\exp\{\alpha S_{n+1}\}\right] \\ = \;& 1 \end{aligned}$$

where the final step follows from (29). Hence:

$$E\left[Y_{n+1}|S_1,\ldots,S_n\right] = \exp\{-\alpha U_n\} = Y_n,$$

and so the proof is complete.

We can now use this fact to find a bound for $\psi(U)$. To do so, let us introduce the positive constant $b > U$, and define the **stopping time** T by:

$$T = \min(n\colon U_n \leq 0 \text{ or } U_n \geq b):$$

Thus, the process stops when the first of two events occurs: (i) ruin; or, (ii) the surplus reaches at least level b. The optional stopping theorem tells us that:

$$E\left[\exp\{-\alpha U_T\}\right] = \exp\{-\alpha U\}.$$

Let p_b denote the probability that ruin occurs without the surplus ever having been at level b or above. Then, conditioning on the two events described above:

$$\begin{aligned} E\left[\exp\{-\alpha U_T\}\right] &= E\left[\exp\{-\alpha U_T\}|U_T \leq 0\right] p_b \\ &\quad + E\left[\exp\{-\alpha U_T\}|U_T \geq b\right](1 - p_b) \\ &= \exp\{-\alpha U\}. \end{aligned} \tag{31}$$

Now let $b \to \infty$. Then $p_b \to \psi(u)$, the first expectation on the right hand side of (31) is at least 1 since it is the moment generating function of the deficit at ruin, evaluated at α, and the second expectation goes to 0 since it is bounded above by $\exp\{-\alpha b\}$. Thus:

$$\exp\{-\alpha U\} = \psi(U) E\left[\exp\{-\alpha U_T\}|U_T \leq 0\right]$$

giving:

$$\psi(U) = \frac{\exp\{-\alpha U\}}{E\left[\exp\{-\alpha U_T\}|U_T \leq 0\right]} \leq \exp\{-\alpha U\}. \tag{32}$$

This gives us a simple bound on the probability of ultimate ruin. It also suggests an appropriate value for the parameter α in formula (29). For example, if the initial surplus is 10, and we require that the probability of ultimate ruin is to be no more than 1%, then we require α such that $\exp\{-10\alpha\} = 0.01$, giving $\alpha = 0.461$.

Let us consider the special case when $\{S_n\}_{n=1}^{\infty}$ is a sequence of independent and identically distributed random variables, each distributed as compound Poisson with Poisson parameter λ. Then P_n is independent of n, say $P_n = P$, and:

$$\begin{aligned} P &= \alpha^{-1} \log E\left[\exp\{\alpha S_n\}\right] \\ &= \alpha^{-1} \log\left[\exp\{\lambda(M_X(\alpha) - 1)\}\right] \end{aligned}$$

where $M_X(.)$ is the moment generating function of the distribution of a single claim amount, giving:

$$\lambda + P\alpha = \lambda M_X(\alpha). \tag{33}$$

This is the equation for the adjustment coefficient for this model (see, for example, Gerber (1979)). Thus, in this particular case, the parameter α is the adjustment coefficient, and equation (32) is simply Lundberg's inequality.

6.5 *Multi-period Analysis — Continuous Time*

Formula (32) shows that when the annual premium is calculated according to formula (29), the probability of ultimate ruin in discrete time for our surplus process is bounded above by $\exp\{-\alpha U\}$. If, in addition, the distribution of aggregate claims each year is normal (assumption (24)) then formula (29) implies that the premium loading is proportional to the variance of the aggregate claims.

We can gain a little more insight, particularly into formula (32), by moving from a discrete time model to a continuous time model. In this section we assume that the aggregate claims in the time interval $[0,t]$ are a random variable $\mu t + \sigma B(t)$, where the stochastic process $\{B(t)\}_{t\geq 0}$ is standard Brownian motion (see Appendix A). This means that in any year, the aggregate claims are distributed as $N(\mu, \sigma^2)$, and are independent of the claims in any other year. This is the continuous time version of assumption (24), but note that we are now assuming that the mean and variance of the annual aggregate claims do not change over time. Brownian motion has stationary and independent increments so for any $0 \leq s < t$:

$$B(t) - B(s) \sim N\left((t-s)\mu, (t-s)\sigma^2\right). \tag{34}$$

We assume that premiums are received continuously at constant rate P *per annum*, where:

$$P = \mu + \frac{1}{2}\alpha\sigma^2 \tag{35}$$

which is equivalent to formula (30). The surplus at time t is denoted $U(t)$, where:

$$U(t) = U + Pt - (\mu t + \sigma B(t)). \tag{36}$$

Now let:

$$\psi_c(U) = \mathrm{P}[U(t) \leq 0 \quad \text{for some } t,\ t > 0]$$

so that $\psi_c(U)$ is the probability of ultimate ruin in continuous time for our surplus process. We will apply the martingale argument of the previous subsection to find $\psi_c(U)$. First note that:

$$\mathrm{E}[\exp\{-\alpha((P-\mu)t - \sigma B(t))\}] = 1. \tag{37}$$

This follows from formulae (34) and (35) and the formula for the moment generating function of the normal distribution.

Next, let $Y(t) = \mathrm{E}[\exp\{-\alpha U(t)\}]$. Then the process $\{Y(t)\}_{t\geq 0}$ is a martingale with respect to[11] $\{B(t)\}_{t\geq 0}$. (Since $\{U(t)\}_{t\geq 0}$ is a continuous time stochastic process, $\{Y(t)\}_{t\geq 0}$ is a martingale in continuous time.) This follows since for $t > s$:

[11] Recall that a martingale is defined with respect to a filtration: here we mean that the relevant filtration is that generated by the process $\{B(t)\}_{t\geq 0}$.

$$\begin{aligned}
& \mathrm{E}[Y(t)|B(u), 0 \leq u \leq s] \\
= \; & \mathrm{E}[\exp\{-\alpha U(t)\}|B(u), 0 \leq u \leq s] \\
= \; & \mathrm{E}[\exp\{-\alpha(U(t)-U(s))\}\exp\{-\alpha U(s)\}|B(u), 0 \leq u \leq s].
\end{aligned}$$

Hence:

$$\begin{aligned}
& \mathrm{E}[Y(t)|B(u), 0 \leq u \leq s] \\
= \; & \mathrm{E}[\exp\{-\alpha(U(t)-U(s))\}|B(u), 0 \leq u \leq s]\mathrm{E}[\exp\{-\alpha U(s)\}|B(u), 0 \leq u \leq s] \\
= \; & \mathrm{E}[\exp\{-\alpha(U(t)-U(s))\}]\exp\{-\alpha U(s)\}.
\end{aligned}$$

Now:

$$U(t)-U(s) = (P-\mu)(t-s) - \sigma(B(t)-B(s)).$$

We can then use the fact that the process $\{B(t)\}_{t\geq 0}$ has stationary increments to say that $B(t) - B(s)$ is equivalent in distribution to $B(t-s)$, and hence $U(t)-U(s)$ is equivalent in distribution to $U(t-s)-U$. (All that stationarity implies is that the distribution of the increment of a process over a given time interval depends only on the length of that time interval, and not on its location. In our context, we are simply interested in the increment of the process $\{B(t)\}_{t\geq 0}$ in a time interval of length $t-s$.) Hence:

$$\begin{aligned}
& \mathrm{E}[\exp\{-\alpha(U(t)-U(s))\}] \\
= \; & \mathrm{E}[\exp\{-\alpha(U(t-s)-U)\}] \\
= \; & E[\exp\{-\alpha((P-\mu)(t-s)+\sigma B(t-s))\}] \\
= \; & 1
\end{aligned}$$

where the final step follows from (37). Hence:

$$\mathrm{E}[Y(t)|B(u), 0 \leq u \leq s] = \exp\{-\alpha U(s)\} = Y(s)$$

and so $\{Y(t)\}_{t\geq 0}$ is a martingale with respect to $\{B(t)\}_{t\geq 0}$.

The optional stopping theorem also applies to martingales in continuous time, so we can use the same argument as in the previous subsection. We define:

$$T_c = \inf\{t\colon U(t) \leq 0 \text{ or } U(t) \geq b)$$

where $b > U$. From the optional stopping theorem we have:

$$E[\exp\{-\alpha U(T_c)\}] = \exp\{-\alpha U\}.$$

Once again defining p_b to be the probability that ruin occurs without the surplus process ever having been at level b or above, we have:

$$\begin{aligned} E\left[\exp\{-\alpha U(T_c)\}\right] &= E\left[\exp\{-\alpha U(T_c)\}|U(T_c) \leq 0\right] p_b \\ &\quad + E\left[\exp\{-\alpha U(T_c)\}|U(T_c) \geq b\right](1-p_b) \\ &= \exp\{-\alpha U\}. \end{aligned}$$

If the surplus level attains b without ruin occurring, then $U(T_c) = b$ since the sample paths of Brownian motion are continuous, i.e. the process cannot jump from below b to above b without passing through b. The situation is the same if ruin occurs. Hence:

$$E\left[\exp\{-\alpha U(T_c)\}|U(T_c) \geq b\right] = \exp\{-\alpha b\}$$

and:

$$E\left[\exp\{-\alpha U(T_c)\}|U(T_c) \leq 0\right] = 1.$$

Thus:

$$p_b + \exp\{-\alpha b\}(1-p_b) = \exp\{-\alpha U\}$$

and if we let $b \to \infty$, then $p_b \to \psi_c(U)$, and hence:

$$\psi_c(U) = \exp\{-\alpha U\}. \tag{38}$$

Formula (38) is Lundberg's inequality for our continuous time surplus process, but this is now an equality. Going back to formula (32), this shows that the upper bound for the probability of ruin in discrete time, in the special case where the mean and variance of claims do not change over time, is just the exact probability of ruin in continuous time.

REFERENCES

BAXTER, M., & RENNIE, A. (1996). *Financial calculus*. Cambridge University Press.

BOWERS, N.L., GERBER, H.U., HICKMAN, J.C., JONES, D.A. & NESBITT, C.J. (1997). *Actuarial mathematics (2nd edition)*. Society of Actuaries, Schaumburg, IL.

GERBER, H.U. (1979). *An introduction to mathematical risk Ttheory*. S.S. Heubner Foundation Monograph Series No. 8. Distributed by R. Irwin, Homewood, IL.

GOOVAERTS, M.J., DE VYLDER, F. & HAEZENDONCK, J. (1984). *Insurance premiums*. North-Holland, Amsterdam.

GRIMMETT, G.R., STIRZAKER, D.R. (1992). *Probability and random processes*. Oxford University Press.

HOEM, J.M. (1969). Markov chain models in life insurance. *Blätter der Deutschen Gesellschaft für Versicherungsmathematik* **9**, 91–107.

HOEM, J.M. (1988). The versatility of the Markov chain as a tool in the mathematics of life insurance. *Transactions of the 23rd International Congress of Actuaries, Helsinki* **S**, 171–202.

HOEM, J.M. & AALEN, O.O. (1978). Actuarial values of payment streams. *Scandinavian Actuarial Journal* **1978**, 38–47.

KLUGMAN, S.A., PANJER, H.H. & WILLMOT, G.E. (1997). *Loss models — from data to decisions*. John Wiley & Sons, New York.

KULKARNI, V.G. (1995). *Modeling and analysis of stochastic systems*. Chapman & Hall, London.

MACDONALD, A.S. (1996a). An actuarial survey of statistical models for decrement and transition data I: Multiple state, Poisson and Binomial models. *British Actuarial Journal* **2**, 129–155.

MACDONALD, A.S. (1996b). An actuarial survey of statistical models for decrement and transition data III: Counting process models. *British Actuarial Journal* **2**, 703–726.

MØLLER, T. (1998). Risk-minimising hedging strategies for unit-linked life insurance contracts. *ASTIN Bulletin* **28**, 17–47.

NEILL, A. (1977). *Life contingencies*. Heinemann, London.

NORBERG, R. (1990). Payment measures, interest and discounting. *Scandinavian Actuarial Journal* **1990**, 14–33.

NORBERG, R. (1991). Reserves in life and pension insurance. *Scandinavian Actuarial Journal* **1991**, 3–24.

NORBERG, R. (1992). Hattendorff's theorem and Theile's differential equation generalized. *Scandinavian Actuarial Journal* **1992**, 2–14.

NORBERG, R. (1993a). Addendum to Hattendorff's theorem and Theile's differential equation generalized, SAJ 1992, 2–14. *Scandinavian Actuarial Journal* **1993**, 50–53.

NORBERG, R. (1993b). Identities for life insurance benefits. *Scandinavian Actuarial Journal* **1993**, 100–106.

NORBERG, R. (1995a). Differential equations for moments of present values in life insurance. *Insurance: Mathematics & Economics* **17**, 171–180.

NORBERG, R. (1995b). Stochastic calculus in actuarial science. *University of Copenhagen Working Paper* **135**, 23 pp.

ØKSENDAL, B. (1998). *Stochastic Differential Equations, 5th Edition*. Springer-Verlag.Berlin

RAMLAU-HANSEN, H. (1988). Hattendorff's theorem: A Markov chain and counting process approach. *Scandinavian Actuarial Journal* **1988**, 143–156.

WILLIAMS, D. (1991). *Probability with martingales*. Cambridge University Press.

WOLTHUIS, H. (1992). Prospective and retrospective premium reserves. *Blätter der Deutschen Gesellschaft für Versicherungsmathematik* **20**, 317–327.

WOLTHUIS, H., & HOEM, J.M. (1990). The retrospective premium reserve. *Insurance: Mathematics & Economics* **9**, 229–234.

APPENDIX A

BROWNIAN MOTION

Suppose that Z_t is a standard Brownian motion under a measure P. Then we have the following properties of Z_t:

(a) Z_t has continuous sample paths which are nowhere differentiable

(b) $Z_0 = 0$

(c) Z_t is normally distributed with mean 0 and variance t

(d) For $0 < s < t$, $Z_t - Z_s$ is normally distributed with mean 0 and variance $t - s$ and it is independent of Z_s

(e) Z_t can be written as the stochastic integral $\int_0^t dZ_s$ where dZ_s can be taken as the increment in Z_t over the small interval $(s, s+ds]$, is normally distributed with mean 0 and variance ds and is independent of Z_s

APPENDIX B

STOCHASTIC DIFFERENTIAL EQUATIONS

A diffusion process, X_t, is a stochastic process which, locally, looks like a scaled Brownian motion with drift. Its dynamics are determined by a stochastic differential equation:

$$dX_t = m(t, X_t)dt + s(t, X_t)dZ_t$$

and we can write down the solution to this as:

$$X_t = X_0 + \int_0^t m(u, X_u)du + \int_0^t s(u, X_u)dZ_u.$$

With traditional calculus we have no problem in dealing with the first integral. However in the second integral the usual Riemann-Stieltjes approach fails because Z_u is just too volatile a function. (This is related to the fact that Z_t is not differentiable.) In fact the second integral is dealt with using Itô integration. A good treatment of this can be found in Øksendal (1998).

Writing down the stochastic integral is not really very informative and it is useful to have, if possible, a closed expression for X_t. An important result which allows us to do this in many cases is **Itô's Lemma**: suppose that X_t and Y_t are diffusion processes with $dX_t = m(t, X_t)dt + s(t, X_t)dZ_t$ and $Y_t = f(t, X_t)$ for some function $f(t, x)$. Then:

$$dY_t = \frac{\partial f}{\partial t}(t, X_t)dt + \frac{\partial f}{\partial x}(t, X_t)dX_t + \frac{1}{2}\frac{\partial^2 f}{\partial x^2}(t, X_t)s(t, X_t)^2 dt$$

For example suppose that $X_t = Z_t$ and $Y_t = \exp[at + bX_t] = f(t, X_t)$. Then:

$$\begin{aligned}
\frac{\partial f}{\partial t} &= a\exp[at + bx] = af(t, x) \\
\frac{\partial f}{\partial x} &= bf(t, x) \\
\frac{\partial^2 f}{\partial x^2} &= b^2 f(t, x).
\end{aligned}$$

Thus, by Itô's Lemma:

$$\begin{aligned} dY_t &= aY_t dt + bY_t dX_t + \frac{1}{2} b^2 Y_t dt \\ &= (a + \frac{1}{2} b^2) Y_t dt + bY_t dZ_t. \end{aligned}$$