Risk and Stochastics in Life Insurance

Ragnar Norberg

Version: January 19, 2006

Contents

Chapter 1

Introduction

1.1 Banking versus insurance

A. The bank savings contract. Upon celebrating his 55th anniversary Mr. (55) (let us call him so) decides to invest money to secure himself economically in his old age. The first idea that occurs to him is to deposit a capital of $S_0 = 1$ (e.g. one hundred thousand pounds) on a savings account today and draw the entire amount with earned compound interest in 15 years, on his 70th birthday. The account bears interest at rate $i = 0.045$ (4.5%) per year. In one year the capital will increase to $S_1 = S_0 + S_0\, i = S_0(1+i)$, in two years it will increase to $S_2 = S_1 + S_1\, i = S_0(1+i)^2$, and so on until in 15 years it will have accumulated to

$$S_{15} = S_0\,(1+i)^{15} = 1.045^{15} = 1.935\,. \tag{1.1}$$

This simple calculation takes no account of the fact that (55) will die sooner or later, maybe sooner than 15 years. Suppose he has no heirs (or he dislikes the ones he has) so that in the event of death before 70 he would consider his savings wasted. Checking population statistics he learns that about 75% of those who are 55 will survive to 70. Thus, the relevant prospects of the contract are:
– with probability 0.75 (55) survives to 70 and will then possess S_{15};
– with probability 0.25 (55) dies before 70 and loses the capital.
In this perspective the expected amount at (55)'s disposal after 15 years is

$$0.75\, S_{15}\,. \tag{1.2}$$

B. A small scale mutual fund. Having thought things over, (55) seeks to make the following mutual arrangement with $(55)^*$ and $(55)^{**}$, who are also 55 years old and are in exactly the same situation as (55). Each of the three deposits $S_0 = 1$ on the savings account, and those who survive to 70, if any, will then share the total accumulated capital $3\, S_{15}$ equally.

The prospects of this scheme are given in Table 1.1, where $+$ and $-$ signify *survival* and *death*, respectively, L_{70} is the number of survivors at age 70, and

Table 1.1: Possible outcomes of a savings scheme with three participants.

(55)	$(55)^*$	$(55)^{**}$	L_{70}	$3\,S_{15}/L_{70}$	Probability
+	+	+	3	S_{15}	$0.75 \cdot 0.75 \cdot 0.75 = 0.422$
+	+	−	2	$1.5\,S_{15}$	$0.75 \cdot 0.75 \cdot 0.25 = 0.141$
+	−	+	2	$1.5\,S_{15}$	$0.75 \cdot 0.25 \cdot 0.75 = 0.141$
+	−	−	1	$3S_{15}$	$0.75 \cdot 0.25 \cdot 0.25 = 0.047$
−	+	+	2	$1.5\,S_{15}$	$0.25 \cdot 0.75 \cdot 0.75 = 0.141$
−	+	−	1	$3\,S_{15}$	$0.25 \cdot 0.75 \cdot 0.25 = 0.047$
−	−	+	1	$3\,S_{15}$	$0.25 \cdot 0.25 \cdot 0.75 = 0.047$
−	−	−	0	undefined	$0.25 \cdot 0.25 \cdot 0.25 = 0.016$

$3\,S_{15}/L_{70}$ is the amount at disposal per survivor (undefined if $L_{70} = 0$). There are now the following possibilities:
– with probability 0.422 (55) survives to 70 together with $(55)^*$ and $(55)^{**}$ and will then possess S_{15};
– with probability $2 \cdot 0.141 = 0.282$ (55) survives to 70 together with one more survivor and will then possess $1.5\,S_{15}$;
– with probability 0.047 (55) survives to 70 while both $(55)^*$ and $(55)^{**}$ die (may they rest in peace) and he will cash the total savings $3S_{15}$;
– with probability 0.25 (55) dies before 70 and will get nothing.
This scheme is superior to the one described in Paragraph A, with separate individual savings contracts: If (55) survives to 70, which is the only scenario of interest to him, he will cash no less than the amount S_{15} he would cash under the individual scheme, and it is likely that he will get more. As compared with (1.2), the expected amount at (55)'s disposal after 15 years is now

$$0.422 \cdot S_{15} + 0.282 \cdot 1.5 \cdot S_{15} + 0.047 \cdot 3\,S_{15} = 0.985\,S_{15}\,.$$

The point is that under the present scheme the savings of those who die are bequeathed to the survivors. Thus the total savings are retained for the group so that nothing is left to others unless the unlikely thing happens that the whole group goes extinct within the term of the contract. This is essentially the kind of solidarity that unites the members of a pension fund. From the point of view of the group as a whole, the probability that all three participants will die before 70 is only 0.016, which should be compared to the probability 0.25 that (55) will die and lose everything under the individual savings program.

C. A large scale mutual scheme. Inspired by the success of the mutual fund idea already on the small scale of three participants, (55) starts to play with the idea of extending it to a large number of participants. Let us assume that a total number of L_{55} persons, who are in exactly the same situation as (55), agree to join a scheme similar to the one described for the three. Then the total savings after 15 years amount to $L_{55}\, S_{15}$, which yields an individual share equal to

$$\frac{L_{55}\, S_{15}}{L_{70}} \tag{1.3}$$

to each of the L_{70} survivors if $L_{70} > 0$. By the so-called law of large numbers, the proportion of survivors L_{70}/L_{55} tends to the individual survival probability 0.75 as the number of participants L_{55} tends to infinity. Therefore, as the number of participants increases, the individual share per survivor tends to

$$\frac{1}{0.75} S_{15}\,, \tag{1.4}$$

and in the limit (55) is faced with the following situation:
– with probability 0.75 he survives to 70 and gets $\frac{1}{0.75} S_{15}$;
– with probability 0.25 he dies before 70 and gets nothing.
The expected amount at (55)'s disposal after 15 years is

$$0.75 \frac{1}{0.75} S_{15} = S_{15}\,,$$

the same as (1.1). Thus, the bequest mechanism of the mutual scheme has raised (55)'s expectations of future pension to what they would be with the individual savings contract if he were immortal. This is what we could expect since, in an infinitely large scheme, some will survive to 70 for sure and share the total savings. All the money will remain in the scheme and will be redistributed among its members by the lottery mechanism of death and survival.

The fact that L_{70}/L_{55} tends to 0.75 as L_{55} increases, and that (1.3) thus stabilizes at (1.4), is precisely what is meant by saying that "insurance risk is diversifiable". The risk can be eliminated by increasing the size of the portfolio.

1.2 Mortality

A. Life and death in the classical actuarial perspective. Insurance mathematics is widely held to be boring. Hopefully, the present text will not support that prejudice. It must be admitted, however, that actuaries use to cheer themselves up with jokes like: "What is the difference between an English and a Sicilian actuary? Well, the English actuary can predict fairly precisely how many English citizens will die next year. Likewise, the Sicilian actuary can predict how many Sicilians will die next year, but he can tell their names as well." The English actuary is definitely the more typical representative of the actuarial profession since he takes a purely statistical view of mortality. Still he is able to analyze insurance problems adequately since what insurance is essentially about, is to average out the randomness associated with the individual risks.

Contemporary life insurance is based on the paradigm of the large scheme (diversification) studied in Paragraph 1.1C. The typical insurance company serves tens and some even hundreds of thousands of customers, sufficiently many to ensure that the survival rates are stable as assumed in Paragraph 1.1C. On the basis of statistical investigations the actuary constructs a so-called *decrement series*, which takes as it starting point a large number ℓ_0 of new-born and, for each age $x = 1, 2, \ldots$, specifies the number of survivors, ℓ_x.

Table 1.2: Excerpt from the mortality table G82M

x:	0	25	50	60	70	80	90
ℓ_x:	100 000	98 083	91 119	82 339	65 024	37 167	9 783
d_x:	58	119	617	1 275	2 345	3 111	1 845
q_x:	.000579	.001206	.006774	.015484	.036069	.083711	.188617
p_x:	.999421	.998794	.993226	.984516	.963931	.916289	.811383

Table 1.2 is an excerpt of the table used by Danish insurers to describe the mortality of insured Danish males. The second row in the table lists some entries of the decrement series. It shows e.g. that about 65% of all new-born will celebrate their 70th anniversary. The number of survivors decreases with age:

$$\ell_x \geq \ell_{x+1} \,.$$

The difference

$$d_x = \ell_x - \ell_{x+1}$$

is the number of deaths at age x (more precisely, between age x and age $x+1$). These numbers are shown in the third row of the table. It is seen that the number of deaths peaks somewhere around age 80. From this it cannot be concluded that 80 is the "most dangerous age". The actuary measures the mortality at any age x by the *one-year mortality rate*

$$q_x = \frac{d_x}{\ell_x} \,,$$

which tells how big proportion of those who survive to age x will die within one year. This rate, shown in the fourth row of the table, increases with the age. For instance, 8.4% of the 80 years old will die within a year, whereas 18.9% of the 90 years old will die within a year. The bottom row shows the one year survival rates

$$p_x = \frac{\ell_{x+1}}{\ell_x} = 1 - q_x \,.$$

We shall present some typical forms of products that an insurance company can offer to (55) and see how they compare with the corresponding arrangements, if any, that (55) can make with his bank.

1.3 Banking

A. Interest. Being unable to find his perfect matches $(55)^*$, $(55)^{**}$,..., our hero (55) abandons the idea of creating a mutual fund and resumes discussions with his bank.

The bank operates with annual interest rate i_t in year $t = 1, 2, ...$ Thus, a unit $S_0 = 1$ deposited at time 0 will accumulate with compound interest as follows: In one year the capital increases to $S_1 = S_0 + S_0\, i_1 = 1 + i_1$, in two years it increases to $S_2 = S_1 + S_1\, i_2 = (1+i_1)(1+i_2)$, and in t years it increases to

$$S_t = (1 + i_1) \cdots (1 + i_t)\,, \tag{1.5}$$

called the t-year *accumulation factor*.

Accordingly, the present value at time 0 of a unit withdrawn in j years is

$$S_j^{-1} = \frac{1}{S_j}\,, \tag{1.6}$$

called the j-year *discount factor* since it is what the bank would pay you at time 0 if you sell to it (discount) a default-free claim of 1 at time j.

Similarly, the value at time t of a unit deposited at time $j < t$ is

$$(1 + i_{j+1}) \cdots (1 + i_t) = \frac{S_t}{S_j}\,,$$

called the accumulation factor over the time period from j to t, and the value at time t of a unit withdrawn at time $j > t$ is

$$\frac{1}{(1 + i_{t+1}) \cdots (1 + i_j)} = \frac{S_t}{S_j}\,,$$

the discount factor over the time period from t to j.

In general, the value at time t of a unit due at time j is $S_t\, S_j^{-1}$, an accumulation factor if $j < t$ and a discount factor if $j > t$ (and of course 1 if $j = t$).

From (1.5) it follows that $S_t = S_{t-1}(1 + i_t)$, hence

$$i_t = \frac{S_t - S_{t-1}}{S_{t-1}}\,,$$

which expresses the interest rate in year t as the relative increase of the balance in year t.

B. Saving in the bank. A general savings contract over n years specifies that at each time $t = 0, ..., n$ (55) is to deposit an amount c_t (*contribution*) and withdraw an amount b_t (*benefit*). The net amount of deposit less withdrawal at time t is $c_t - b_t$. At any time t the cash balance of the account, henceforth also called *the retrospective reserve*, is the total of past (including present) deposits less withdrawals compounded with interest,

$$U_t = S_t \sum_{j=0}^{t} S_j{}^{-1}(c_j - b_j)\,. \tag{1.7}$$

It develops in accordance with the “forward” recursive scheme

$$U_t = U_{t-1}(1 + i_t) + c_t - b_t\,, \tag{1.8}$$

$t = 1, 2, \ldots, n$, commencing from

$$U_0 = c_0 - b_0\,.$$

Each year (55) will receive from the bank a statement of account with the calculation (1.8), showing how the current balance emerges from the previous balance, the interest earned meanwhile, and the current movement (deposit less withdrawal).

The balance of a savings account must always be non-negative,

$$U_t \geq 0\,, \tag{1.9}$$

and at time n, when the contract terminates and the account is closed, it must be null,

$$U_n = 0\,. \tag{1.10}$$

In the course of the contract the bank must maintain a so-called *prospective reserve* to meet its future liabilities to the customer. At any time t the adequate reserve is

$$V_t = S_t \sum_{j=t+1}^{n} S_j{}^{-1}(b_j - c_j)\,, \tag{1.11}$$

the present value of future withdrawals less deposits . Similar to (1.8), the prospective reserve is calculated by the “backward” recursive scheme

$$V_t = (1 + i_{t+1})^{-1}(b_{t+1} - c_{t+1} + V_{t+1})\,, \tag{1.12}$$

$t = n-1, n-2, \ldots, 0$, starting from

$$V_n = 0\,.$$

The constraint (1.10) is equivalent to

$$\sum_{j=0}^{n} S_j{}^{-1}c_j - \sum_{j=0}^{n} S_j{}^{-1}b_j = 0\,, \tag{1.13}$$

which says that the discounted value of deposits must be equal to the discounted value of the withdrawals. It implies that, at any time t, the retrospective reserve equals the prospective reserve,

$$U_t = V_t \, ,$$

as is easily verified. (Insert the defining expression (1.7) with $t = n$ into (1.10), split the sum $\sum_{j=0}^{n}$ into $\sum_{j=0}^{t} + \sum_{j=t+1}^{n}$, and multiply with S_t/S_n, to arrive at $U_t - V_t = 0$.)

C. The endowment contract. The bank proposes a savings contract according to which (55) saves a fixed amount c annually in 15 years, at ages 55,...,69, and thereafter withdraws $b = 1$ (one hundred thousand pounds, say) at age 70. Suppose the annual interest rate is fixed and equal to $i = 0.045$, so that the accumulation factor in t years is $S_t = (1+i)^t$, the discount factor in j years is ${S_j}^{-1} = (1+i)^{-j}$, and the time t value of a unit due at time j is

$$S_t \, {S_j}^{-1} = (1+i)^{t-j} \, .$$

For the present contract the equivalence requirement (1.13) is

$$\sum_{j=0}^{14} (1+i)^{-j} c \; - \; (1+i)^{-15} 1 \; = \; 0 \, ,$$

from which the bank determines

$$c \; = \; \frac{(1+i)^{-15}}{\sum_{j=0}^{14} (1+i)^{-j}} \; = \; 0.04604 \, , \tag{1.14}$$

Due to interest, this amount is considerably smaller than $1/15 = 0.06667$, which is what (55) would have to save per year if he should choose to tuck the money away under his mattress.

1.4 Insurance

A. The life endowment. Still, to (55) 0.04604 (four thousand six hundred and four pounds) is a considerable expense. He believes in a life before death, and it should be blessed with the joys that money can buy. He talks to an insurance agent, and is delighted to learn that, under a life annuity policy designed precisely as the savings scheme, he would have to deposit an annual amount of only 0.03743 (three thousand seven hundred and forty three pounds).

The insurance agent explains: The calculations of the bank depend only on the amounts $c_t - b_t$ and would apply to any customer (x) who would enter into the same contract at age x, say. Thus, to the bank the customer is really an unknown Mr. X. To the insurance company, however, he is not just Mr. X, but the significant Mr. (x) now x years old. Working under the hypothesis that (x) is one of the ℓ_x survivors at age x in the decrement series and that they all hold identical contracts, the insurer offers (x) a general life annuity policy whereby each deposit or withdrawal is conditional on survival. For the entire portfolio the retrospective reserve at time t is

$$\begin{aligned} U_t^p &= S_t \sum_{j=0}^{t} S_j^{-1}(c_j - b_j)\,\ell_{x+j} & (1.15)\\ &= U_{t-1}^p(1+i_t) + (c_t - b_t)\,\ell_{x+t}\,. & (1.16)\end{aligned}$$

The prospective portfolio reserve at time t is

$$\begin{aligned} V_t^p &= S_t \sum_{j=t+1}^{n} S_j^{-1}(b_j - c_j)\ell_{x+j} & (1.17)\\ &= (1+i_{t+1})^{-1}((b_{t+1} - c_{t+1})\ell_{x+t+1} + V_{t+1}^p)\,. & (1.18)\end{aligned}$$

In particular, for the life endowment analogue to (55)'s savings contract, the only payments are $c_t = c$ for $t = 0, \ldots, 14$ and $b_{15} = 1$. The equivalence requirement (1.13) becomes

$$\sum_{j=0}^{14}(1+i)^{-j} c\,\ell_{55+j} \;-\; (1+i)^{-15} 1\,\ell_{70} \;=\; 0\,, \tag{1.19}$$

from which the insurer determines

$$c \;=\; \frac{(1+i)^{-15}\,\ell_{70}}{\sum_{j=0}^{14}(1+i)^{-j}\ell_{55+j}} \;=\; 0.03743\,. \tag{1.20}$$

Inspection of the expressions in (1.14) and (1.20) shows that the latter is smaller due to the fact that ℓ_x is decreasing. This phenomenon is known as *mortality bequest* since the savings of the deceased are bequeathed to the survivors. We shall pursue this issue in Paragraph C below.

B. A life assurance contract. Suppose, contrary to the former hypothesis, that (55) has dependents whom he cares for. Then he might be concerned that, if he should die within the term of the contract, the survivors in the pension scheme will be his heirs, leaving his wife and kids with nothing. He figures that, in the event of his untimely death before the age of 70, the family would need a down payment of $b = 1$ (one hundred thousand pounds) to compensate the loss of their bread-winner. The bank can not help in this matter; the benefit of b would have to be raised immediately since (55) could die tomorrow, and it would be meaningless to borrow the money since full repayment of the loan would be due immediately upon death. The insurance company, however, can offer (55) a so-called term life assurance policy that provides the wanted death benefit against an affordable annual premium of $c = 0.01701$.

The equivalence requirement (1.13) now becomes

$$\sum_{j=0}^{14}(1+i)^{-j} c\,\ell_{55+j} \;-\; \sum_{j=1}^{15}(1+i)^{-j}\,1\,d_{55+j-1} \;=\; 0\,, \tag{1.21}$$

from which the insurer determines

$$c = \frac{\sum_{j=1}^{15}(1+i)^{-j}\, d_{55+j-1}}{\sum_{j=0}^{14}(1+i)^{-j}\ell_{55+j}} = 0.01701\,. \tag{1.22}$$

C. Individual reserves and mortality bequest. In the insurance schemes described above the contracts of deceased members are void, and the reserves of the portfolio are therefore to be shared equally between the survivors at any time. Thus, we introduce the individual retrospective and prospective reserves at time t,

$$U_t = U_t^p/\ell_{x+t}\,, \qquad V_t = V_t^p/\ell_{x+t}\,.$$

Since we have established that $U_t = V_t$, we shall henceforth be referring to them as the individual reserve or just the reserve.

For the general pension insurance contract in Paragraph A we get from (1.16) that the individual reserve develops as

$$\begin{aligned} U_t &= U_{t-1}\frac{\ell_{x+t-1}}{\ell_{x+t}}(1+i_t) + (c_t - b_t) \\ &= U_{t-1}\left(1 + \frac{d_{x+t-1}}{l_{x+t}}\right)(1+i_t) + (c_t - b_t)\,. \end{aligned} \tag{1.23}$$

The bequest mechanism is clearly seen by comparing (1.23) to (1.8): the additional term $U_{t-1}(1+i_t)\, d_{x+t-1}/l_{x+t}$ in the latter is precisely the share per survivor of the savings left over to them by those who died during the year. Virtually, the mortality bequest acts as an increase of the interest rate.

Table 1.3 shows how the reserve develops for the endowment contracts offered by the bank and the insurance company, respectively. It is seen that the insurance scheme requires a smaller reserve than the bank savings scheme.

Table 1.3: Reserve $U_t = V_t$ for bank savings account and for life endowment insurance

t :	0	4	9	14
Savings account:	0.04604	0.25188	0.56577	0.95694
Life endowment:	0.03743	0.21008	0.49812	0.92523

For the life assurance described in Paragraph B we obtain similarly that the individual reserve develops as as shown in Table 1.4.

D. Insurance risk in a finite portfolio. The perfect balance in (1.19) and (1.21) rests on the hypothesis that the decrement series ℓ_{x+t} follows the pattern of an infinitely large portfolio. In a finite portfolio, however, the factual numbers of survivors, L_{x+t}, will be subject to randomness and will be determined by

Table 1.4: Reserve $U_t = V_t$ for a term life assurance of 1 against level premium in 15 years from age 55

t :	0	4	9	14
	0.01701	0.04460	0.06010	0.03170

the survival probabilities p_{x+t} (some of which are) shown in Table 1.2. The difference between discounted premiums and discounted benefits,

$$D = \sum_{j=0}^{14} (1+i)^{-j} 0.0374\, L_{55+j} \; - \; (1+i)^{-15}\, L_{70}\,,$$

will be a random quantity. It will have expected value 0, and its standard deviation measures how much insurance risk is left due to "imperfect diversification" in a finite portfolio. An easy exercise in probability calculus shows that the standard deviation of D/L_{55} is $\frac{1}{\sqrt{L_{55}}}0.1685$. It tends to 0 as L_{55} goes to infinity.

For the term insurance contract the corresponding quantity is $\frac{1}{\sqrt{L_{55}}}0.3478$, indicating that term insurance is a more risky business than life endowment.

1.5 With-profit contracts: Surplus and bonus

A. With-profit contracts. Insurance policies are long term contracts, with time horizons wide enough to capture significant variations in interest and mortality. For simplicity we shall focus on interest rate uncertainty and assume that the mortality law remains unchanged over the term of the contract. We will discuss the issue of surplus and bonus in the framework of the life endowment contract considered in Paragraph 1.4.A.

At time 0, when the contract is written with benefits and premiums binding to both parties, the future development of the interest rates i_t is uncertain, and it is impossible to foresee what premium level c will establish the required equivalence

$$\sum_{j=0}^{14} {S_j}^{-1} c\, \ell_{55+j} = {S_{15}}^{-1} 1\, \ell_{70}\,, \tag{1.24}$$

with

$$S_j = (1+i_1)\cdots(1+i_j)\,.$$

If it should turn out that, due to adverse development of interest and mortality, premiums are insufficient to cover the benefit, then there is no way the insurance company can avoid a loss; it cannot reduce the benefit and it cannot increase the premiums since these were irrevocably set out in the contract at time 0. The only way the insurance company can prevent such a loss, is to charge a premium 'on the safe side', high enough to be adequate under all likely scenarios. Then, if everything goes well, a surplus will accumulate. This surplus belongs to the insured and is to be repaid as so-called *bonus*, e.g. as increased benefits or reduced premiums.

B. First order basis. The usual way of setting premiums to the safe side is to base the calculation of the premium level and the reserves on a provisional *first order basis*, assuming a fixed annual interest rate i^*, which represents a worst case scenario and leads to higher premium and reserves than are likely to be needed. The corresponding accumulation factor is $S_t^* = (1+i^*)^t$. The individual reserve based on the prudent first order assumptions is called the *first order reserve*, and we denote it by V_t^* as before. The premiums are determined so as to satisfy equivalence under the first order assumption.

C. Surplus. At any time t we define the technical surplus Q_t as the difference between the retrospective reserve under the factual interest development and the retrospective reserve under the first order assumption:

$$\begin{aligned} Q_t &= S_t \sum_{j=0}^{t} S_j^{-1}\, c\, \ell_{55+j} - S_t^* \sum_{j=0}^{t} S_j^{*-1}\, c\, \ell_{55+j} \\ &= S_t \sum_{j=0}^{t-1} S_j^{-1}\, c\, \ell_{55+j} - S_t^* \sum_{j=0}^{t-1} S_j^{*-1}\, c\, \ell_{55+j}\,. \end{aligned}$$

Setting $S_t = S_{t-1}(1+i_t)$ and $S_t^* = S_{t-1}^*(1+i^*)$, writing $1+i^* = 1+i_t-(i_t-i^*)$ in the latter, and rearranging a bit, we find that Q_t develops as

$$Q_t = Q_{t-1}(1+i_t) + V_{t-1}^*\,(i_t - i^*)\,\ell_{55+t-1}\,, \tag{1.25}$$

commencing from

$$Q_0 = 0\,.$$

The contribution to the technical surplus in year t is

$$V_{t-1}^*\,(i_t - i^*)\,\ell_{55+t-1}\,,$$

which is easy to interpret: it is precisely the interest earned on the reserve in excess of what has been assumed under the prudent first order assumption.

The surplus is to be redistributed as *bonus*. Several bonus schemes are used in practice. One can repay currently the contribution $V_{t-1}^*\,(i_t - i^*)\,\ell_{55+t-1}$ as so-called *cash bonus* (a premium deductible), whereby each survivor at time t will receive

$$V_{t-1}^*\,(i_t - i^*)\,\ell_{55+t-1}/\ell_{55+t}\,.$$

Another possibility is to postpone repayment until the term of the contracts and grant so-called *terminal bonus* to the survivors (an added benefit), the amount per survivor being b^+ given by

$$S_{15} \sum_{j=1}^{15} S_j^{-1} V^*_{j-1} \, (i_j - i^*) \, \ell_{55+j} \; = \; \ell_{70} b^+ \, .$$

Between these two solutions there are countless other possibilities. In any case, the point is that the financial risk can be eliminated: the insurer observes the development of the factual interest and only in arrears repays the insured so as to restore equivalence on the basis of the factual interest rate development. This works well provided the first order interest rate is set on the safe side so that $i_t \geq i^*$ for all t.

There is a problem, however: Negative bonus can never be applied. Therefore the insurer will suffer a loss if the factual interest falls short of the technical interest rate. In this perspective cash bonus is the most risky solution and terminal bonus is the least risky solution.

If the financial market is sufficiently rich in assets, then the interest rate guarantee that is thus inherent in the with-profit policy can be priced, and the insured can be charged an extra premium to cover it. This would ultimately eliminate the financial risk by diversifying, not only the insurance portfolio, but also the investment portfolio.

1.6 Unit-linked insurance

A quite different way of going about the financial risk is the so-called unit-linked contract. As the name indicates, the idea is to relate payments directly to the development of the investment portfolio, i.e. the interest rate. Consider the balance equation for an endowment of b against premium c_t in year $t = 1, \ldots, 14$:

$$S_{15} \sum_{t=0}^{14} S_t^{-1} \, c_t \, \ell_{55+t} \; - \; b \, \ell_{70} \; = \; 0 \, . \tag{1.26}$$

A perfect link between payments and investments performance is obtained by letting the premiums and the benefit be inflated by the index S,

$$c_t = S_t \, c \, ,$$

and

$$b_{15} = S_{15} \, 1 \, .$$

Here c is a baseline premium, which is to be determined. Then (1.26) becomes

$$S_{15} \sum_{t=0}^{14} S_t^{-1} \, S_t \, c \, \ell_{55+t} \; - \; S_{15} \, \ell_{70} \; = \; 0 \, ,$$

which reduces to

$$\sum_{t=0}^{14} c \, \ell_{55+t} \; - \; \ell_{70} \; = \; 0 \, ,$$

and we find

$$c = \frac{\ell_{70}}{\sum_{t=0}^{14} \ell_{55+t}}\,.$$

Again financial risk has been perfectly eliminated and diversification of the insurance portfolio is sufficient to establish balance between premiums and benefits.

Perfect linking as defined here is not common in practice. Presumably, remnants of social welfare thinking have led insurers to modify the unit-linked concept in various ways, typically by introducing a guarantee on the sum insured to the effect that it cannot be less than 1 (say). Also the premium is usually not index-linked. Under such modified variations of the unit-linked policy one cannot in general obtain balance by the simple device above. However, the problem can in principle be resolved by calculating the price of the financial claim thus introduced and to charge the insured with the needed additional premium.

1.7 Issues for further study

The simple pieces of actuarial reasoning in the previous sections involve two constituents, interest and mortality, and these are to be studied separately in the two following chapters. Next we shall escalate the discussion to more complex situations. For instance, suppose (55) wants a life insurance that is paid out only if his wife survives him, or with a sum insured that depends on the number of children that are still alive at the time of his death. Or he may demand a pension payable during disability or unemployment. We need also to study the risk associated with insurance, which is due to the uncertain developments of the insurance portfolio and the investment portfolio: the deaths in a finite insurance portfolio do not follow the mortality table (1.2) exactly, and the interest earned on the investments may differ from the assumed 4.5% per year, and neither can be predicted precisely at the outset when the policies are issued.

In a scheme of the classical mutual type the problem was how to share existing money in a fair manner. A typical insurance contract of today, however, specifies that certain benefits will be paid contingent on certain events related only to the individual insured under the contract. An insurance company working with this concept in a finite portfolio, with imperfect diversification of insurance risk, faces a risk of insolvency as indicated in Paragraph 1.3.D. In addition comes the financial risk, and ways of getting around that were indicated in Sections 1.5 and 1.6. The total risk has to be controlled in some way. With these issues in mind, we now commence our studies of the theory of life insurance.

The reader is advised to consult the following authoritative textbooks on the subject: [6] (a good classic – sharpen your German!), [4], [29] (lexicographic, treats virtually every variation of standard insurance products, and includes a good chapter on population theory), [46] (an excellent early text based on probabilistic models, placing emphasis on risk considerations), [11], [15] (an original approach to the field – sharpen your French!), and [23] (the most recent of the mentioned texts, still classical in its orientation).

Chapter 2

Payment streams and interest

2.1 Basic definitions and relationships

A. Streams of payments. What is money? In lack of a precise definition you may add up the face values of the coins and notes you find in your purse and say that the total amount is your money. Now, if you do this each time you open your purse, you will realize that the development of the amount over time is important. In the context of insurance and finance the time aspect is essential since payments are usually regulated by a contract valid over some period of time. We will give precise mathematical content to the notion of payment streams and, referring to Appendix A, we deal only with their properties as functions of time and do not venture to discuss their possible stochastic properties for the time being.

To fix ideas and terminology, consider a financial contract commencing at time 0 and terminating at a later time n ($\leq \infty$), say, and denote by A_t the total amount paid in respect of the contract during the time interval $[0, t]$. The *payment function* $\{A_t\}_{t \geq 0}$ is assumed to be the difference of two non-decreasing, finite-valued functions representing incomes and outgoes, respectively, and is thus of finite variation (FV). Furthermore, the payment function is assumed to be right-continuous (RC). From a practical point of view this assumption is just a matter of convention, stating that the balance of the account changes at the time of any deposit or withdrawal. From a mathematical point of view it is convenient, since payment functions can then serve as integrators. In fact, we shall restrict attention to payment functions that are piece-wise differentiable (PD):

$$A_t = A_0 + \int_0^t a_\tau \, d\tau + \sum_{0 < \tau \leq t} \Delta A_\tau \,, \tag{2.1}$$

where $\Delta A_\tau = A_\tau - A_{\tau-}$. The integral adds up payments that fall due continuously, and the sum adds up lump sum payments. In differential form (2.1) reads

$$dA_t = a_t\, dt + \Delta A_t\,. \tag{2.2}$$

It seems natural to count incomes as positive and outgoes as negative. Sometimes, and in particular in the context of insurance, it is convenient to work with outgoes less incomes, and to avoid ugly minus signs we introduce $B = -A$.

Having explained what payments are, let us now see how they accumulate under the force of interest. There are monographs written especially for actuaries on the topic, see [31] and [17], but we will gather the basics of the theory in only a few lines.

B. The time value of money: interest and discounting. Suppose money is currently invested on (or borrowed from) an account that bears interest. This means that a unit deposited on the account at time s gives the account holder the right to cash, at any other time t, a certain amount $S(s,t)$, typically different from 1. The function S must be strictly positive, and we shall argue that it must satisfy the functional relationship

$$S(s,u) = S(s,t)\, S(t,u)\,, \tag{2.3}$$

implying, of course, that $S(t,t) = 1$ (put $s = t = u$ and use strict positivity): If the account holder invests 1 at time s, he may cash the amount on the left of (2.3) at time u. If he instead withdraws the value $S(s,t)$ at time t and immediately reinvests it again, he will obtain the amount on the right of (2.3) at time u. To avoid arbitrary gains, so-called *arbitrage*, the two strategies must give the same result.

From (2.3) it follows that, for any fixed t and u,

$$S(t,u) = \frac{S(s,u)}{S(s,t)}$$

holds true for all s. In particular, choosing $s = 0$ and introducing $S_t = S(0,t)$, we obtain

$$S(t,u) = \frac{S_u}{S_t}\,, \tag{2.4}$$

which says that the value at time u of a unit invested at time t is the ratio of the values at times u and t of a unit invested at time 0. We have actually proved that a function $S(t,u)$ satisfies (2.3) if and only if it is of the form (2.4) for some strictly positive function S_t such that

$$S_0 = 1\,.$$

(The proof above if for the “only if” part, and the “if” part is trivial.) The value S_t at time t of a unit deposited at time 0, will be called the *accumulation* *function*. Correspondingly, S_t^{-1} is the value at time 0 of a unit withdrawn at time t, and we call it the *discount function*.

We will henceforth assume that S_t is of the form

$$S_t = e^{\int_0^t r} \,, \qquad S_t^{-1} = e^{-\int_0^t r} \,, \tag{2.5}$$

where r_t is some piece-wise continuous function, usually positive. (The short-hand exemplified by $\int r = \int r_\tau \, d\tau$ will be in frequent use throughout.) Accumulation factors of this form are invariably used in basic banking operations (loans and savings) and also for bonds issued by governments and corporations.

Under the rule (2.5) the dynamics of accumulation and discounting are given by

$$dS_t = S_t \, r_t \, dt \,, \tag{2.6}$$

$$dS_t^{-1} = -S_t^{-1} \, r_t \, dt \,. \tag{2.7}$$

The relation (2.6) says that the interest earned in a small time interval is proportional to the length of the interval and to the current amount on deposit. The proportionality factor r_t is called the *force of interest* or the *(instantaneous) interest rate* at time t. In integral form (2.6) reads

$$S_t = S_s + \int_s^t S_\tau \, r_\tau \, d\tau \,, \qquad s \le t \,, \tag{2.8}$$

and (2.7) reads $S_u^{-1} = S_t^{-1} - \int_t^u S_\tau^{-1} \, r_\tau \, d\tau$ or

$$S_t^{-1} = S_u^{-1} + \int_t^u S_\tau^{-1} \, r_\tau \, d\tau \,, \qquad t \le u \,. \tag{2.9}$$

We will be working with the expressions

$$S(u,t) = e^{-\int_t^u r}$$

for the general *discount factor* when $t \le u$ and

$$S(s,t) = e^{\int_s^t r}$$

for the general *accumulation factor* when $t \ge s$.

By constant interest rate r we have $S_t = e^{rt}$ and and $S_t^{-1} = e^{-rt}$. Upon introducing the the *annual interest rate*

$$i = e^r - 1 \,, \tag{2.10}$$

whereby the *annual accumulation factor* is $S_1 = (1+i)$, and the *annual discount factor*

$$v = e^{-r} = \frac{1}{1+i} \,, \tag{2.11}$$

we have

$$S_t = (1+i)^t \,, \qquad S_t^{-1} = v^t \,. \tag{2.12}$$

C. Valuation of payment streams. Suppose that the incomes/outgoes created by the payment stream A are currently deposited on/drawn from an account which bears interest at rate r_t at time t. By (2.4) the value at time t of the amount dA_τ paid in the small time interval around time τ is $e^{\int_0^t r} e^{-\int_0^\tau r}\, dA_\tau$. Summing over all time intervals we get the value at time t of the entire payment stream,

$$e^{\int_0^t r} \int_{0-}^{n} e^{-\int_0^\tau r}\, dA_\tau = U_t - V_t\,,$$

where

$$U_t = e^{\int_0^t r} \int_{0-}^{t} e^{-\int_0^\tau r}\, dA_\tau = \int_{0-}^{t} e^{\int_\tau^t r}\, dA_\tau \tag{2.13}$$

is the accumulated value of past incomes less outgoes, and (recall the convention $B = -A$)

$$V_t = e^{\int_0^t r} \int_{t}^{n} e^{-\int_0^\tau r}\, dB_\tau = \int_{t}^{n} e^{-\int_t^\tau r}\, dB_\tau \tag{2.14}$$

is the discounted value of future outgoes less incomes. This decomposition is particularly relevant for payments governed by some contract; U_t is the *cash balance*, that is, the amount held at the time of consideration, and V_t is the future liability. The difference between the two is the current value of the contract.

The development of the cash balance can be viewed in various ways: Application of (A.8) to (2.13), taking $X_t = e^{\int_0^t r}$ (continuous, with dynamics given by (2.6)) and $Y_t = \int_{0-}^{t} e^{-\int_0^\tau r}\, dA_\tau$, yields

$$dU_t = U_t r_t\, dt + dA_t\,, \tag{2.15}$$

By definition,

$$U_0 = A_0\,. \tag{2.16}$$

Integrating (2.15) from 0 to t, using the initial condition (2.16), gives

$$U_t = A_t + \int_0^t U_\tau r_\tau\, d\tau\,. \tag{2.17}$$

An alternative expression,

$$U_t = A_t + \int_0^t e^{\int_\tau^t r} A_\tau\, r_\tau\, d\tau\,, \tag{2.18}$$

is derived from (2.13) upon applying the rule (A.9) of integration by parts:

$$\begin{aligned}
\int_{0-}^{t} e^{-\int_0^\tau r}\, dA_\tau &= A_0 + \int_0^t e^{-\int_0^\tau r}\, dA_\tau \\
&= A_0 + e^{-\int_0^t r} A_t - A_0 - \int_0^t A_\tau e^{-\int_0^\tau r}(-r_\tau)\, d\tau\,.
\end{aligned}$$

The relationships (2.15) – (2.18) show how the cash balance emerges from payments and earned interest. They are easy to interpret and can be read aloud in non-mathematical terms.

It follows from (2.18) that, if $r \geq 0$, then an increase of A results in an increase of U. In particular, advancing payments of a given amount produces a bigger cash balance.

Likewise, from (2.14) we derive

$$dV_t = V_t r_t \, dt - dB_t \,. \tag{2.19}$$

By definition, if $n < \infty$,

$$V_n = 0 \,, \tag{2.20}$$

or, setting $\Delta B_n = B_n - B_{n-}$,

$$V_{n-} = \Delta B_n \,. \tag{2.21}$$

Integrating (2.19) from t to n, using the ultimo condition (2.20), gives the following analogue to (2.17):

$$V_t = B_n - B_t - \int_t^n V_\tau \, r_\tau \, d\tau \,. \tag{2.22}$$

The analogue to (2.18) is

$$V_t = B_n - B_t - \int_t^n e^{-\int_t^\tau r} (B_n - B_\tau) \, r_\tau \, d\tau \,. \tag{2.23}$$

The last two relationships are valid for $n = \infty$ only if $B_\infty < \infty$. Yet another expression is

$$V_t = e^{-\int_t^n r} (B_n - B_t) + \int_t^n e^{-\int_t^\tau r} (B_\tau - B_t) \, r_\tau \, d\tau \,, \tag{2.24}$$

which is obtained upon integrating by parts in (2.23) or, simpler, multiplying $B_n - B_t$ with

$$1 = e^{-\int_t^n r} + \int_t^n e^{-\int_t^\tau r} \, r_\tau \, d\tau$$

(a twist on (2.9)) and gathering terms.

Again interpretations are easy. The relations (2.22) and (2.23) state, in different ways, that the debt can be settled immediately at a price which is the total debt minus the present value of future interest saved by advancing the repayment. The relation (2.24) states that repayment can be postponed until the term of the contract at the expense of paying interest on the outstanding amounts meanwhile. It follows from (2.24) that, if $r \geq 0$, then an increase of the outstanding payments produces an increase in the reserve. In particular, advancing payments of a given amount leads to a bigger reserve.

Typically, the financial contract will lay down that incomes and outgoes be equivalent in the sense that

$$U_n = 0 \quad \text{or} \quad V_{0-} = 0\,. \tag{2.25}$$

These two relationships are equivalent and they imply that, for any t,

$$U_t = V_t\,. \tag{2.26}$$

We anticipate here that, in the insurance context, the equivalence requirement is usually not exercised at the level of the individual policy: the very purpose of insurance is to redistribute money among the insured. Thus the principle must be applied at the level of the portfolio in some sense, which we shall discuss later. Moreover, in insurance the payments, and typically also the interest rate, are not known at the outset, so in order to establish equivalence one may have to currently adapt the payments to the development in some way or other.

D. Some standard payment functions and their values. Certain simple payment functions are so frequently used that they have been given names. An *endowment* of 1 at time n is defined by $A_t = \varepsilon_n(t)$, where

$$\varepsilon_n(t) = \begin{cases} 0, & 0 \le t < n\,, \\ 1, & t \ge n\,. \end{cases} \tag{2.27}$$

(The only payment is $\Delta A_n = 1$.) By constant interest rate r the present value at time 0 of the endowment is e^{-rn} or, recalling the notation in Chapter 1, v^n.

An n-year *immediate annuity* of 1 per year consists of a sequence of endowments of 1 at times $t = 1, \ldots, n$, and is thus given by

$$A_t = \sum_{j=1}^{n} \varepsilon_j(t) = [t] \wedge n\,.$$

By constant interest rate r its present value at time 0 is

$$a_{\overline{n}|} = \sum_{j=1}^{n} e^{-rj} = \frac{1 - e^{-rn}}{i}\,, \tag{2.28}$$

see (2.11) – (2.10).

An n-year *annuity-due* of 1 per year consists of a sequence of endowments of 1 at times $t = 0, \ldots, n-1$, that is,

$$A_t = \sum_{j=0}^{n-1} \varepsilon_j(t) = [t+1] \wedge n\,.$$

By constant interest rate its present value at time 0 is

$$\ddot{a}_{\overline{n}|} = \sum_{j=0}^{n-1} e^{-rj} = (1+i)\, a_{\overline{n}|} = (1+i)\frac{1 - e^{-rn}}{i}\,. \tag{2.29}$$

An n-year *continuous annuity* payable at level rate 1 per year is given by

$$A_t = t \wedge n \, . \tag{2.30}$$

For the case with constant interest rate its present value at time 0 is (recall (2.11))

$$\bar{a}_{\overline{n|}} = \int_0^n e^{-r\tau} \, d\tau = \frac{1 - e^{-rn}}{r} \, . \tag{2.31}$$

An everlasting (perpetual) annuity is called a *perpetuity*. Putting $n = \infty$ in the (2.28), (2.29), and (2.31), we find the following expressions for the present values of the immediate perpetuity, the perpetuity-due, and the continuous perpetuity:

$$a_{\overline{\infty|}} = \frac{1}{i} \, , \quad \ddot{a}_{\overline{\infty|}} = \frac{1+i}{i} \, , \quad \bar{a}_{\overline{\infty|}} = \frac{1}{r} \, . \tag{2.32}$$

An m-year *deferred* n-year temporary life annuity commences only after m years and is payable throughout n years thereafter. Thus it is just the difference between an $m + n$ year annuity and an m year annuity. For the continuous version,

$$A_t = ((t - m) \vee 0) \wedge n = (t \wedge (m + n)) - (t \wedge m) \, . \tag{2.33}$$

Its present value at time 0 by constant interest is denoted $\bar{a}_{m|n}$ and must be

$$\bar{a}_{m|n} = \bar{a}_{\overline{m+n|}} - \bar{a}_{\overline{m|}} = v^m \bar{a}_{\overline{n|}} \, . \tag{2.34}$$

2.2 Application to loans

A. Basic features of a loan contract. Loans and saving accounts in banks are particularly simple financial contracts for which interest is invariably calculated in accordance with (2.5). Let us consider a loan contract stipulating that at time 0, say, the bank pays to the borrower an amount H, called the *principal* ('first' in Latin), and that the borrower thereafter pays back or *amortizes* the loan in accordance with a non-decreasing payment function $\{A_t\}_{0 \le t \le n}$ called the *amortization function*. The term of the contract, n, is sometimes called the duration of the loan. Without loss of generality we assume henceforth that $H = 1$ (the principal is proclaimed monetary unit).

The amortization function is to fulfill $A_0 = 0$ and $A_n \ge 1$. The excess of total amortizations over the principal is the total amount of *interest*. We denote it by R_n and have $A_n = 1 + R_n$. General principles of book-keeping, needed e.g. for taxation purposes, prescribe that the decomposition of the amortizations into repayments and interest be extended to all $t \in [0, n]$. Thus,

$$A_t = F_t + R_t \, , \tag{2.35}$$

where F is a non-decreasing *repayment function* satisfying

$$F_0 = 0\,, \qquad F_n = 1$$

(formally a distribution function due to the convention $H = 1$), and R is a non-decreasing *interest payment function*.

Furthermore, the contract is required to specify a *nominal force of interest* r_t, $0 \leq t \leq n$, under which the value of the amortizations should be equivalent to the value of the principal, that is,

$$\int_0^n e^{-\int_0^\tau r}\, dA_\tau = 1\,. \tag{2.36}$$

There are, of course, infinitely many admissible decompositions (2.35) satisfying (2.36). A clue to constraints on F and R is offered by the relationship

$$\int_0^n e^{-\int_0^\tau r}\, dR_\tau = \int_0^n e^{-\int_0^\tau r}(1 - F_\tau)\, r_\tau\, d\tau\,, \tag{2.37}$$

which is obtained upon inserting (2.35) into (2.36) and then using integration by parts on the term $\int_0^n \exp\left(-\int_0^\tau r\right) dF_\tau = -\int_0^n \exp\left(-\int_0^\tau r\right) d(1 - F_\tau)$. The condition (2.37) is trivially satisfied if

$$dR_t = (1 - F_t)\, r_t\, dt\,,$$

that is, interest is paid currently and instantaneously on the *outstanding (part of the) principal*, $1 - F$. This will be referred to as *natural interest*.

Under the scheme of natural interest the relation (2.35) becomes

$$dA_t = dF_t + (1 - F_t)\, r_t\, dt\,, \tag{2.38}$$

which establishes a one-to-one correspondence between amortizations and repayments. The differential equation (2.38) is easily solved:

First, integrate (2.38) over $(0, t]$ to obtain

$$A_t = F_t + \int_0^t (1 - F_\tau)\, r_\tau\, d\tau\,, \tag{2.39}$$

which determines amortizations when repayments are given.

Second, multiply (2.38) with $\exp\left(-\int_0^t r\right)$ to obtain $\exp\left(-\int_0^t r\right) dA_t = -d\left(\exp\left(-\int_0^t r\right)(1 - F_t)\right)$ and then integrate over $(t, n]$ to arrive at

$$\int_t^n e^{-\int_t^\tau r} dA_\tau = 1 - F_t\,, \tag{2.40}$$

which determines (outstanding) repayments when amortizations are given.

The relationships (2.39) and (2.40) are easy to interpret. For instance, since $1 - F_t$ is the remaining debt at time t, (2.40) is the time t update of the equivalence requirement (2.36). When it comes to numerical computation, the integral expressions in (2.39) and (2.40) are not so useful, however. Whether we want to compute A for given F or the other way around, we would use the differential equation (2.38).

B. Standard forms of loans. We list some standard types of loans, taking now r constant. It is understood that we consider only times t in $[0, n]$.

The simplest form is the *fixed loan*, which is repaid in its entirety only at the term of the contract, that is, $F_t = \varepsilon_n(t)$, the endowment defined by (2.27). The amortization function is obtained directly from (2.39): $A_t = \varepsilon_n(t) + rt$.

A *series loan* has repayments of annuity form. The continuous version is given by $F_t = t/n$, see (2.30). The amortization plan is obtained from (2.39): $A_t = t/n + rt(1 - t/2n)$. Thus, $dF_t/dt = 1/n$ (fixed) and $dR_t/dt = r(1 - t/n)$ (linearly decreasing).

An *annuity loan* is called so because the amortizations, which are the amounts actually paid by the borrower, are of annuity form. The continuous version is given by $A_t = t/\bar{a}_{\overline{n|}}$, see (2.36) and (2.31). From (2.40) we easily obtain $F_t = 1 - \bar{a}_{\overline{n-t|}}/\bar{a}_{\overline{n|}}$. We find $dF_t/dt = e^{-r(n-t)}/\bar{a}_{\overline{n|}}$ (exponentially increasing), and $dR_t/dt = (1 - e^{-r(n-t)})/\bar{a}_{\overline{n|}}$.

Putting $n = \infty$, the fixed loan and the series loan both specialize to an infinite loan without repayment. Amortizations consist only of interest, which is paid indefinitely at rate r.

Chapter 3

Mortality

3.1 Aggregate mortality

A. The stochastic model. Consider an aggregate of individuals, e.g. the population of a nation, the persons covered under an insurance scheme, or a certain species of animals. The individuals need not be animate beings; for instance, in engineering applications one is often interested in studying the work-life until failure of technical components or systems. Having demographic and actuarial problems in mind, we shall, however, be speaking of persons and life lengths until death.

Due to differences in inheritance and living conditions and also due to events of a more or less purely random nature, like accidents, diseases, etc., the life lengths vary among individuals. Therefore, the life length of a randomly selected new-born can suitably be represented by a non-negative random variable T with a cumulative distribution function

$$F(t) = \mathbb{P}[T \leq t]\,. \tag{3.1}$$

In survival analysis it is convenient to work with the *survival function*

$$\bar{F}(t) = \mathbb{P}[T > t] = 1 - F(t)\,. \tag{3.2}$$

Fig. 3.1 shows F and $\bar{F}$ for the mortality law G82M used by Danish life insurers as a basis for calculating premiums for insurances on male lives. Find the median life length and some other percentiles of this life distribution by inspection of the graphs!

We assume that F is absolutely continuous and denote the density by f;

$$f(t) = \frac{d}{dt}F(t) = -\frac{d}{dt}\bar{F}(t)\,. \tag{3.3}$$

The density of the distribution in Fig. 3.1 is depicted in Fig. 3.2. Find the mode by inspection of the graph! Can you already at this stage figure why the median and the mode of F in Fig. 3.1 appear to exceed those of the mortality law of the Danish male population?

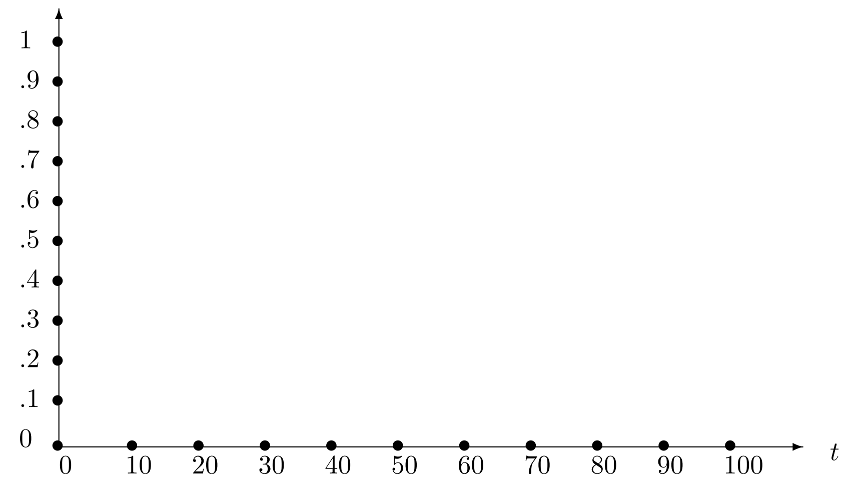


Figure 3.1: The G82M mortality law: F broken line, $\bar{F}$ whole line.

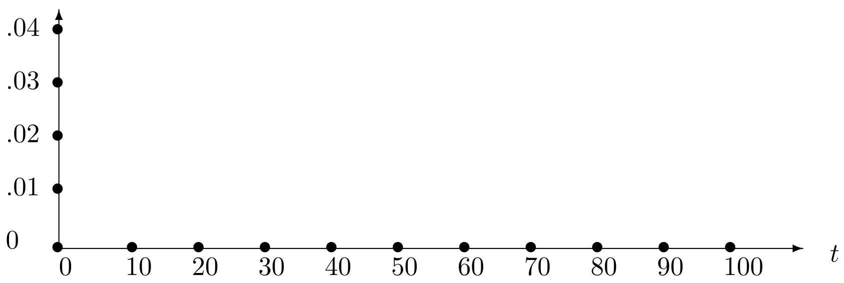


Figure 3.2: The density f for the G82M mortality law.

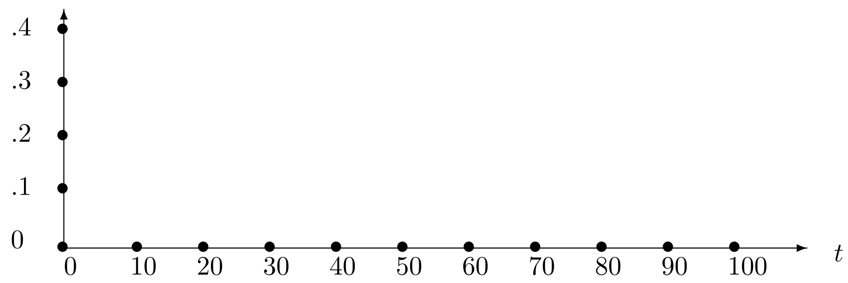


Figure 3.3: The force of mortality μ for the G82M mortality law.

B. The force of mortality. The density is the derivative of $-\bar{F}$, see (3.3). When dealing with non-negative random variables representing life lengths, it is convenient to work with the derivative of $-\ln \bar{F}$,

$$\mu(t) = \frac{d}{dt}\{-\ln \bar{F}(t)\} = \frac{f(t)}{\bar{F}(t)}\,, \tag{3.4}$$

which is well defined for all t such that $\bar{F}(t) > 0$. For small, positive dt we have

$$\mu(t)dt = \frac{f(t)dt}{\bar{F}(t)} = \frac{\mathbb{P}[t < T \leq t + dt]}{\mathbb{P}[T > t]} = \mathbb{P}[T \leq t + dt \,|\, T > t]\,.$$

(In the second equality we have neglected a term $o(dt)$ such that $o(dt)/dt \to 0$ as $dt \searrow 0$.) Thus, for a person aged t, the probability of dying within dt years is (approximately) proportional to the length of the time interval, dt. The proportionality factor $\mu(t)$ depends on the attained age, and is called *the force of mortality* at age t. It is also called the *mortality intensity* or *hazard rate* at age t, the latter expression stemming from reliability theory, which is concerned with the durability of technical devices.

Fig 3.3 shows the force of mortality corresponding to F in Fig. 3.1. Assess roughly the probability that a t years old person will die within one year for $t = 60,\ 70,\ 80,\ 90$!

Integrating (3.4) from 0 to t and using $\bar{F}(0) = 1$, we obtain

$$\bar{F}(t) = e^{-\int_0^t \mu}. \tag{3.5}$$

Relation (3.4) may be cast as

$$f(t) = \bar{F}(t)\mu(t) = e^{-\int_0^t \mu}\mu(t)\,, \tag{3.6}$$

which says that the probability $f(t)dt$ of dying in the age interval $(t, t+dt)$ is the product of the probability $\bar{F}(t)$ of survival to t and the conditional probability $\mu(t)dt$ of then dying before age $t + dt$.

The functions F, $\bar{F}$, f, and μ are equivalent representations of the mortality law; each of them corresponds one-to-one to any one of the others.

Since $\bar{F}(\infty) = 0$, we must have $\int_0^\infty \mu = \infty$. Thus, if there is a finite highest attainable age ω such that $\bar{F}(\omega) = 0$ and $\bar{F}(t) > 0$ for $t < \omega$, then $\int_0^t \mu \nearrow \infty$ as $t \nearrow \omega$. If, moreover, μ is non-decreasing, we must also have $\lim_{t \nearrow \omega} \mu(t) = \infty$.

C. The distribution of the remaining life length. Let T_x denote the remaining life length of an individual chosen at random from the x years old members of the population. Then T_x is distributed as $T - x$, conditional on $T > x$, and has cumulative distribution function

$$F(t|x) = \mathbb{P}[T \leq x + t \,|\, T > x] = \frac{F(x+t) - F(x)}{1 - F(x)}$$

and survival function

$$\bar{F}(t|x) = \mathbb{P}[T > x + t \,|\, T > x] = \frac{\bar{F}(x+t)}{\bar{F}(x)}\,, \tag{3.7}$$

which are well defined for all x such that $\bar{F}(x) > 0$. The density of this conditional distribution is

$$f(t|x) = \frac{f(x+t)}{\bar{F}(x)}\,. \tag{3.8}$$

Denote by $\mu(t|x)$ the force of mortality of the distribution $F(t|x)$. It is obtained by inserting $f(t|x)$ from (3.8) and $\bar{F}(t|x)$ from (3.7) in the places of f and $\bar{F}$ in the definition (3.4). We find

$$\mu(t\,|\,x) = f(x+t)/\bar{F}(x+t) = \mu(x+t)\,. \tag{3.9}$$

Alternatively, we could insert (3.5) into (3.7) to obtain

$$\bar{F}(t|x) = e^{-\int_x^{x+t} \mu(y)\, dy} = e^{-\int_0^t \mu(x+\tau) d\tau}\,, \tag{3.10}$$

which by the general relation (3.5) entails (3.9). Relation (3.9) explains why the force of mortality is particularly handy; it depends only on the attained age $x + t$, whereas the conditional density in (3.8) depends in general on x and t in a more complex manner. Thus, the properties of all the conditional survival distributions are summarized by one simple function of the total age only.

Figs. 3.4 – 3.6 depict the functions $F(t|70)$, $\bar{F}(t|70)$, $f(t|70)$, and $\mu(t|70) = \mu(70 + t)$ derived from the life time distribution in Fig. 3.1. Observe that the first three of these functions are obtained simply by scaling up the corresponding graphs in Figs. 3.1 – 3.2 by the factor $1/\bar{F}(70)$ over the interval $(70, \infty)$. The force of mortality remains unchanged, however.

D. Expected values in life distributions. Let T be a non-negative random variable with absolutely continuous distribution function F, and let $G : \mathbb{R}_+ \to \mathbb{R}$ be a PD and RC function such that $\mathbb{E}[G(T)]$ exists and is finite. Integrating by parts in the defining expression

$$\mathbb{E}[G(T)] = \int_0^\infty G(\tau)\, dF(\tau)\,,$$

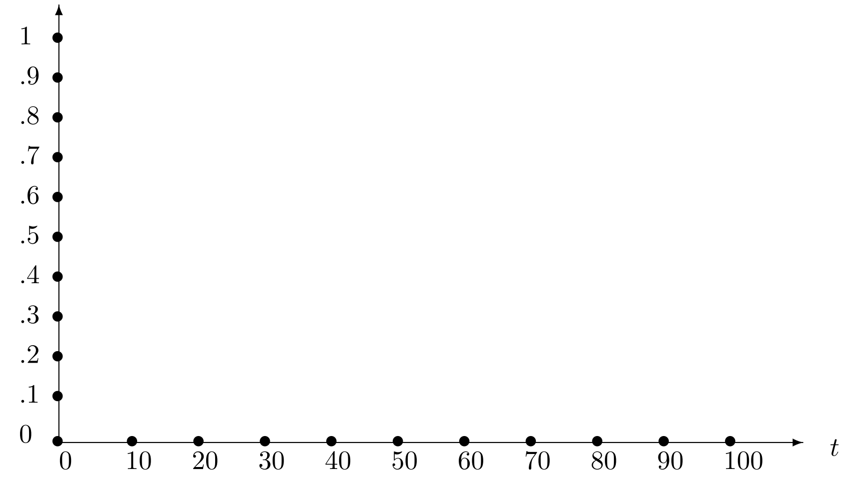


Figure 3.4: Conditional distribution of remaining life length for the G82M mortality law: $F(t|70)$ broken line, $\bar{F}(t|70)$ whole line.

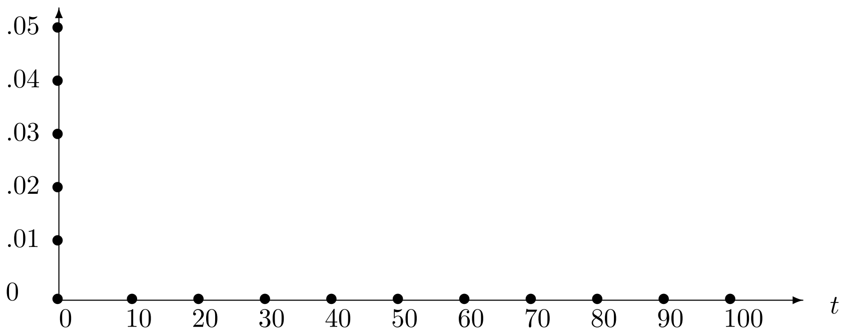


Figure 3.5: Conditional density of remaining life length $f(t|70)$ for the G82M mortality law.

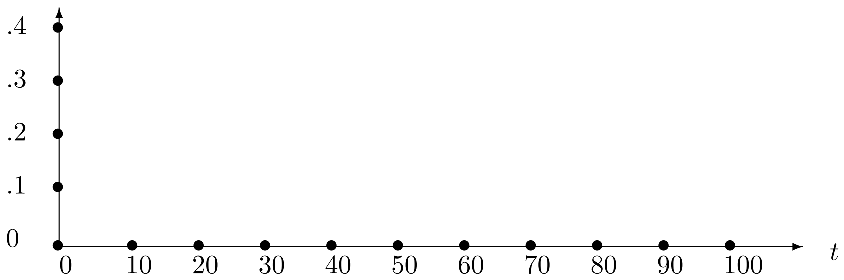


Figure 3.6: The force of mortality $\mu(t|70) = \mu(70 + t)$, $t > 0$, for the G82M mortality law.

we find

$$\mathbb{E}[G(T)] = G(0) + \int_0^\infty \bar{F}(\tau)\, dG(\tau)\,. \tag{3.11}$$

Taking $G(t) = t^k$ we get

$$\mathbb{E}[T^k] = k \int_0^\infty t^{k-1} \bar{F}(t)\, dt\,, \tag{3.12}$$

and, in particular,

$$\mathbb{E}[T] = \int_0^\infty \bar{F}(t)\, dt\,. \tag{3.13}$$

The expected remaining life time for an x years old person is

$$\bar{e}_x = \int_0^\infty \bar{F}(t \,|\, x)\, dt\,. \tag{3.14}$$

From (3.10) it is seen that $\bar{F}(t \,|\, x)$ is a decreasing function of x for fixed t if μ is an increasing function. Then $\bar{e}_x$ is a decreasing function of x. One can easily construct mortality laws for which $\bar{F}(t \,|\, x)$ and $\bar{e}_x$ are not decreasing functions of x.

Consider the more general function

$$G(t) = ((t \wedge b) - (t \wedge a))^k = \begin{cases} 0\,, & 0 \le t < a, \\ (t-a)^k\,, & a \le t < b, \\ (b-a)^k\,, & b \le t, \end{cases} \tag{3.15}$$

that is, $dG(t) = k(t-a)^{k-1}\, dt$ for $a < t < b$ and 0 elsewhere. It is realized that $G(T)$ is the kth power of the number of years lived between age a and age b. From (3.11) we obtain

$$\mathbb{E}[G(T)] = k \int_a^b (t-a)^{k-1} \bar{F}(t)\, dt\,, \tag{3.16}$$

In particular, the expected number of years lived between the ages of a and b is $\int_a^b \bar{F}(t)\, dt$, which is the area between the t-axis and the survival function in the interval from a to b. The formula can be motivated directly by noting that $\bar{F}(t)\, dt$ is the expected number of years survived in the small time interval $(t, t + dt)$ and using that the "expected value of the sum is the sum of the expected values".

3.2 Some standard mortality laws

A. The exponential distribution. Suppose the force of mortality is $\mu(t) = \lambda$, independent of the age. This means there are no wear-out effects; each morning when you wake up (if you wake up) life starts anew with the same prospects of longevity as for a new-born. Then the survival function (3.5) becomes

$$\bar{F}(t) = e^{-\lambda t}\,, \tag{3.17}$$

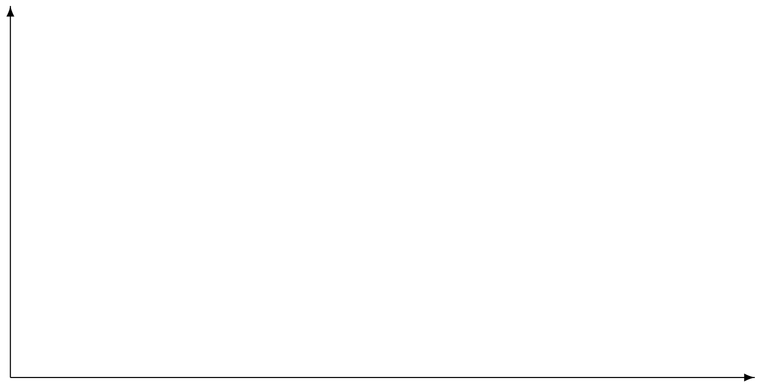


Figure 3.7: Two exponential laws with intensities λ_1 and λ_2 such that $\lambda_1 < \lambda_2$; $\bar{F}_1$ and $\bar{F}_2$ whole line, f_1 and f_2 broken line.

and the density (3.6) becomes

$$f(t) = \lambda e^{-\lambda t} \,. \tag{3.18}$$

Thus, T is exponentially distributed with parameter λ. The conditional survival function (3.10) becomes $\bar{F}(t \,|\, x) = e^{-\lambda t}$, hence

$$\bar{F}(t \,|\, x) = \bar{F}(t) \,, \tag{3.19}$$

the same as (3.17). The exponential distribution is a suitable model for certain technical devices like bulbs and electronic components. Unfortunately, it is not so apt for description of human lives.

One could arrive at the exponential distribution by specifying that (3.19) be valid for all x and t, that is, the probability of surviving another t years is independent of the age x. Then, from the general relation (3.7) we get

$$\bar{F}(x+t) = \bar{F}(x)\bar{F}(t) \tag{3.20}$$

for all non-negative x and t. It follows by induction that for each pair of positive integers m and n, $\bar{F}(\frac{m}{n}) = \bar{F}(\frac{1}{n})^m = \bar{F}(1)^{\frac{m}{n}}$, hence

$$\bar{F}(t) = \bar{F}(1)^t \tag{3.21}$$

for all positive rational t. Since $\bar{F}$ is right-continuous, (3.21) must hold true for all $t > 0$. Putting $\bar{F}(1) = e^{-\lambda}$, we arrive at (3.17).

Fig. 3.7 shows the survival function and the density for two different values of λ.

B. The Weibull distribution. The intensity of this distribution is of the form

$$\mu(t) = \beta\alpha^{-\beta}t^{\beta-1}\,, \tag{3.22}$$

$\alpha, \beta > 0$. The corresponding survival function is $\bar{F}(t) = \exp\left(-(\frac{t}{\alpha})^{\beta}\right)$.

If $\beta > 1$, then $\mu(t)$ is increasing, and if $\beta < 1$, then $\mu(t)$ is decreasing. If $\beta = 1$, the Weibull law reduces to the exponential law. Draw the graphs of $\bar{F}$ and f for some different choices of α and β!

We have $\mu(x+t) = \beta\alpha^{-\beta}(x+t)^{\beta-1}$, and, by virtue of (3.22), $\bar{F}(t\,|\,x)$ is not a Weibull law.

C. The Gompertz-Makeham distribution. This distribution is widely used as a model for survivorship of human lives, especially in the context of life insurance. Thus, as it will be frequently referred to, we shall use the acronym G-M for this law. Its mortality intensity is of the form

$$\mu(t) = \alpha + \beta e^{\gamma t}\,, \tag{3.23}$$

$\alpha, \beta \geq 0$. The corresponding survival function is

$$\bar{F}(t) = \exp\left(-\int_0^t (\alpha + \beta e^{\gamma s})ds\right) = \exp\left(-\alpha t - \beta(e^{\gamma t} - 1)/\gamma\right)\,. \tag{3.24}$$

If $\beta > 0$ and $\gamma > 0$, then $\mu(t)$ is an increasing function of t. The constant term α accounts for age-independent causes of death like certain accidents and epidemic diseases, and the term $\beta e^{\gamma t}$ accounts for all kinds of wear-out effects due to aging.

We have $\mu(x+t) = \alpha + \beta e^{\gamma x}e^{\gamma t}$, and so (3.23) shows that $\bar{F}(t|x)$ is also a G-M survival function with parameters α, $\beta e^{\gamma x}$, γ. The special case $\alpha = 0$ is referred to as the (pure) Gompertz law.

The G82M mortality law depicted in Fig. 3.1 is the G-M law with

$$\alpha = 5 \cdot 10^{-4}\,, \qquad \beta = 7.5858 \cdot 10^{-5}\,, \qquad \gamma = \ln(1.09144)\,. \tag{3.25}$$

Table E.1 in Appendix E shows $\mu(t)$, $\bar{F}(t)$ and $f(t)$ for integer t.

3.3 Actuarial notation

A. Actuaries in all countries – unite! The International Association of Actuaries (IAA) has laid down a standard notation, which is generally accepted among actuaries all over the world. Familiarity with this notation is a must for anyone who wants to communicate in writing or reading with actuaries, and we shall henceforth adopt it in those simple situations where it is applicable.

B. A list of some standard symbols. According to the IAA standard, the quantities introduced so far are denoted as follows:

$$
\begin{aligned}
{}_tq_x &= F(t\,|\,x)\,, && (3.26)\\
{}_tp_x &= \bar{F}(t\,|\,x)\,, && (3.27)\\
\mu_{x+t} &= \mu(x+t)\,. && (3.28)
\end{aligned}
$$

In particular, ${}_tq_0 = F(t)$ and ${}_tp_0 = \bar{F}(t)$. One-year death and survival probabilities are abbreviated as

$$q_x = {}_1q_x\,, \quad p_x = {}_1p_x\,. \tag{3.29}$$

Frequently used is also the "n-year deferred probability of death within m years",

$$_{n|m}q_x = {}_{m+n}q_x - {}_nq_x = {}_np_x - {}_{m+n}p_x = {}_np_x\,{}_mq_{x+n}\,. \tag{3.30}$$

The formulas in Section 3.1 are easily translated, e.g.

$$
\begin{aligned}
{}_tp_x &= \exp(-\int_0^t \mu_{x+\tau}\,d\tau), && (3.31)\\
f(t\,|\,x) &= {}_tp_x\mu_{x+t}, && (3.32)\\
\bar{e}_x &= \int_0^\infty {}_tp_x dt. && (3.33)
\end{aligned}
$$

Frequently actuaries work with expected numbers of survivors instead of probabilities. Consider a population of l_0 new-born who are subject to the same law of mortality given by (3.28). The expected number of survivors at age x is

$$l_x = l_0\,{}_xp_0\,. \tag{3.34}$$

The function $\{l_x;\ x > 0\}$ is called the *decrement function* or, when considered only at integer values of x, the *decrement series.* Expressed in terms of the decrement function we find e.g.

$$
\begin{aligned}
{}_tp_x &= l_{x+t}/l_x\,, && (3.35)\\
\mu_{x+t} &= -l'_{x+t}/l_{x+t}\,, && (3.36)\\
f(t\,|\,x) &= -l'_{x+t}/l_x\,, && (3.37)\\
\bar{e}_x &= \int_0^\infty l_{x+t}\,dt/l_x\,. && (3.38)
\end{aligned}
$$

The pieces of IAA notation we have shown here are quite pleasing to the eye and also space-saving; for instance, the symbol on the left of (3.27) involves three typographical entities, whereas the one on the right involves six.

3.4 Select mortality

A. The insurance portfolio consists of selected lives. Consider an individual who purchases a life insurance at age x. In short, he will be referred to as (x) in what follows.

It is quite common in actuarial practice to assume that the force of mortality of (x) depends on x and t in a more complex manner than the simple relationship (3.9), which rested on the assumption that (x) is chosen at random from the x years old individuals in the population. The fact that (x) purchases insurance adds information to the mere fact that he has attained age x; he does not represent a purely random draw from the population, but is rather selected by some mechanisms. It is easy to think of examples of such mechanisms. For instance that poor people can not afford to buy insurance and, to the extent that longevity depends on economic situation, the mortality experience for insured people would reflect that they are wealthy enough to buy insurance ('survival of the fattest'). Judging from textbooks on life insurance, e.g. [4] and [29] and many others, it seems that the underwriting standards of the insurer are generally held to be the predominant selective mechanism; before an insurance policy is issued, the insurer must be satisfied that the applicant meets certain requirements with regard to health, occupation, and other factors that are assumed to determine the prospects of longevity. Only first class lives are eligible to insurance at ordinary rates.

Thus there is every reason to account for selection effects by letting the force of mortality be some more general function $\mu_x(t)$ or, in other words, specify that T_x follows a survival function $F_x(t)$ that is not necessarily of the form (3.7). One then speaks of *select mortality*.

B. More of actuarial notation. The standard actuarial notation for select mortality is

$$
\begin{aligned}
{}_\tau q_{[x]+t} &= \mathbb{P}[T_x \le t+\tau \,|\, T_x > t]\,, && (3.39)\\
{}_\tau p_{[x]+t} &= \mathbb{P}[T_x > t+\tau \,|\, T_x > t]\,, && (3.40)\\
\mu_{[x]+t} &= \lim_{h \searrow 0} \frac{{}_h q_{[x]+t}}{h}\,. && (3.41)
\end{aligned}
$$

The idea is that the both the current age, $x+t$, and the age at entry, x, are directly visible in $[x]+t$.

From a technical point of view select mortality is just as easy as aggregate mortality; we work with the distribution function ${}_t q_{[x]}$ instead of ${}_t q_x$, and are interested in it as a function of t. For instance,

$$
{}_\tau p_{[x]+t} = \frac{{}_{t+\tau} p_{[x]}}{{}_t p_{[x]}} = \exp\left(-\int_t^{t+\tau} \mu_{[x]+s}\, ds\right)\,.
$$

C. Features of select mortality. There is ample empirical evidence to support the following facts about select mortality in life insurance populations:

– For insured lives of a given age the rate of mortality usually increases with increasing duration.
– The effect of selection tends to decrease with increasing duration and becomes negligible for practical purposes when the duration exceeds a certain *select period*.
– The mortality among insured lives is generally lower than the mortality in the population.

There are many possible ways of building such features into the model. For instance, one could modify the aggregate G-M intensity as

$$\mu_{[x]+t} = \alpha(t) + \beta(t)\, e^{\gamma\,(x+t)}\,,$$

where α and β are non-negative and non-decreasing functions bounded from above. In Section 7.6 we shall show how the selection mechanism can be explained in models that describe more aspects of the individual life histories than just survival and death.

Chapter 4

Insurance of a single life

4.1 Some standard forms of insurance

A. The single-life status. Consider a person aged x with remaining life length T_x as described in the previous section. In actuarial parlance this life is called the *single-life status* (x) . Referring to Appendix B, we introduce the indicator of the event of survival in t years, $I_t = 1[T_x > t]$. This is a binomial random variable with 'success' probability ${}_tp_x$. The indicator of the event of death within t years is $1 - I_t = 1[T_x \leq t]$, which is a binomial variable with 'success' probability ${}_tq_x = 1 - {}_tp_x$. (We apologize for sometimes using technical terms where they may sound misplaced.) Note that, being 0 or 1, any indicator $1[A]$ satisfies $1[A]^q = 1[A]$ for $q > 0$.

The present section lists some standard forms of insurance that (x) can purchase, investigates some of their properties, and presents some basic actuarial methods and formulas.

We assume that the investments of the insurance company yield interest at a fixed rate r so that accumulation and discounting take place in in accordance with (2.12).

B. The pure endowment insurance. An *n-year pure (life) endowment* of 1 is a unit that is paid to (x) at the end of n years if he is then still alive. In other words, the associated payment function is an endowment of I_n at time n. Its present value at time 0 is

$$PV^{e;n} = e^{-rn} I_n \,. \tag{4.1}$$

The expected value of $PV^{e;n}$, denoted by ${}_nE_x$, is

$${}_nE_x = e^{-rn} \, {}_np_x \,. \tag{4.2}$$

For any $q > 0$ we have $(PV^{e;n})^q = e^{-qrn} I_n$ (recall that $I_n^q = I_n$), and so the q-th non-central moment of $PV^{e;n}$ can be expressed as

$$\mathbb{E}[(PV^{e;n})^q] = {}_nE_x^{(qr)} \,, \tag{4.3}$$

where the top-script (qr) signifies that discounting is made under a force of interest that is qr.

In particular, the variance of $PV^{e;n}$ is

$$\mathbb{V}[PV^{e;n}] = {}_nE_x^{(2r)} - {}_nE_x^2 \,. \tag{4.4}$$

C. The life assurance. A life assurance contract specifies that a certain amount, called the *sum insured*, is to be paid upon the death of the insured, possibly limited to a specified period. We shall here consider only insurances payable immediately upon death, and take the sum to be 1 (just a matter of notation).

First, an n-year *term insurance* is payable upon death within n years. The payment function is a lump sum of $1 - I_n$ at time T_x. Its present value at time 0 is

$$PV^{ti;n} \quad = \quad e^{-rT_x}\,(1 - I_n)\,. \tag{4.5}$$

The expected value of $PV^{ti;n}$ is

$$\bar{A}_{\overset{1}{x}\,\overline{n}|} = \int_0^n e^{-r\tau}\,{}_\tau p_x\,\mu_{x+\tau}\,d\tau\,, \tag{4.6}$$

and, similar to (4.3),

$$\mathbb{E}[(PV^{ti;n})^q] = \bar{A}^{(qr)}_{\overset{1}{x}\,\overline{n}|}\,. \tag{4.7}$$

In particular,

$$\mathbb{V}[PV^{ti;n}] = \bar{A}^{(2r)}_{\overset{1}{x}\,\overline{n}|} - \bar{A}^{2}_{\overset{1}{x}\,\overline{n}|}\,. \tag{4.8}$$

An n-year *endowment insurance* is payable upon death if it occurs within time n and otherwise at time n. The payment function is a lump sum of 1 at time $T_x \wedge n$. Its present value at time 0 is

$$PV^{ei;n} = e^{-r(T_x \wedge n)}\,. \tag{4.9}$$

The expected value of $PV^{ei;n}$ is

$$\bar{A}_{x\,\overline{n}|} = \int_0^n e^{-r\tau}\,{}_\tau p_x\,\mu_{x+\tau}\,d\tau + e^{-rn}\,{}_np_x = \bar{A}_{\overset{1}{x}\,\overline{n}|} + {}_nE_x\,, \tag{4.10}$$

and

$$\mathbb{E}(PV^{ei;n})^q = \bar{A}^{(qr)}_{x\,\overline{n}|}\,. \tag{4.11}$$

It follows that

$$\mathbb{V}[PV^{ei;n}] = \bar{A}^{(2r)}_{x\,\overline{n}|} - \bar{A}^{2}_{x\,\overline{n}|}\,. \tag{4.12}$$

D. The life annuity. An *n-year temporary life annuity* of 1 per year is payable as long as (x) survives but limited to n years. We consider here only the continuous version. Recalling (2.30), the associated payment function is an annuity of 1 in $T_x \wedge n$ years. Its present value at time 0 is

$$PV^{a;n} = \bar{a}_{\overline{T_x \wedge n}|} = \frac{1 - e^{-r(T_x \wedge n)}}{r} \,. \tag{4.13}$$

The expected value of $PV^{a;n}$ is

$$\bar{a}_{x\,\overline{n}|} = \int_0^n \bar{a}_{\overline{\tau}|}\, {}_\tau p_x\, \mu_{x+\tau}\, d\tau + \bar{a}_{\overline{n}|}\, {}_n p_x \,.$$

A more appealing formula is

$$\bar{a}_{x\,\overline{n}|} = \int_0^n e^{-r\tau}\, {}_\tau p_x\, d\tau \,, \tag{4.14}$$

which displays the life annuity as a "continuum of life endowments", $\bar{a}_{x\,\overline{n}|} = \int_0^n {}_\tau E_x\, d\tau$. There are several ways of proving (4.14). Using brute force, one can integrate by parts:

$$\begin{aligned} \bar{a}_{\overline{n}|}\, {}_n p_x &= \bar{a}_{\overline{0}|}\, {}_0 p_x + \int_0^n \frac{d}{d\tau} \bar{a}_{\overline{\tau}|}\, {}_\tau p_x\, d\tau + \int_0^n \bar{a}_{\overline{\tau}|} \frac{d}{d\tau}\, {}_\tau p_x\, d\tau \\ &= \int_0^n e^{-r\tau}\, {}_\tau p_x\, d\tau - \int_0^n \bar{a}_{\overline{\tau}|}\, {}_\tau p_x\, \mu_{x+\tau}\, d\tau \,. \end{aligned}$$

Using the brain instead, one realizes that the expected present value at time 0 of the payments in any small time interval $(\tau, \tau + d\tau)$ is $e^{-r\tau}\, d\tau\, {}_\tau p_x$, and summing over all time intervals one arrives at (4.14) ("the expected value of a sum is the sum of the expected values"). This kind of reasoning will be omnipresent throughout the text, and would also immediately produce formula (4.6) and (4.10). The recipe is: *Find the expected present value of the payments in each small time interval and add up.*

We shall demonstrate below that

$$\mathbb{E}[(PV^{a;n})^q] = \frac{q}{r^{q-1}} \sum_{p=1}^{q} (-1)^{p-1} \binom{q-1}{p-1} \bar{a}^{(pr)}_{x\,\overline{n}|} \,, \tag{4.15}$$

from which we derive

$$\mathbb{V}[PV^{a;n}] = \frac{2}{r} \left(\bar{a}_{x\,\overline{n}|} - \bar{a}^{(2r)}_{x\,\overline{n}|} \right) - \bar{a}^2_{x\,\overline{n}|} \,. \tag{4.16}$$

The endowment insurance is a combined benefit consisting of an n-year term insurance and an n-year pure endowment. By (4.9) and (4.13) it is related to the life annuity by

$$PV^{a;n} = \frac{1 - PV^{ei;n}}{r} \quad \text{or} \quad PV^{ei;n} = 1 - rPV^{a;n} \,, \tag{4.17}$$

which just reflects the more general relationship (2.31). Taking expectation in (4.17), we get

$$\bar{A}_{x\,\overline{n}|} = 1 - r\bar{a}_{x\,\overline{n}|}\,. \tag{4.18}$$

Also, since $PV^{ti;n} = PV^{ei;n} - PV^{e;n} = 1 - rPV^{a;n} - PV^{e;n}$, we have

$$\bar{A}_{\overset{1}{x}\,\overline{n}|} = 1 - r\bar{a}_{x\,\overline{n}|} - {}_n E_x\,. \tag{4.19}$$

The formerly announced result (4.15) follows by operating with the q-th moment on the first relationship in (4.17), and then using (4.12) and (4.18) and rearranging a bit. One needs the binomial formula

$$(x+y)^q = \sum_{p=0}^{q} \binom{q}{p} x^{q-p} y^p$$

and the special case $\sum_{p=0}^{q} \binom{q}{p} (-1)^{q-p} = 0$ (for $x = -1$ and $y = 1$).

A whole-life annuity is obtained by putting $n = \infty$. Its expected present value is denoted simply by $\bar{a}_x$ and is obtained by putting $n = \infty$ in (4.14), that is

$$\bar{a}_x = \int_0^\infty e^{-r\tau}\ {}_\tau p_x\, d\tau, \tag{4.20}$$

and the same goes for the variance in (4.16) (justify the limit operations).

E. Deferred benefits. *An m-year deferred n-year temporary life annuity* commences only after m years, provided that (x) is then still alive, and is payable throughout n years thereafter as long as (x) survives. The present value of the benefits is

$$\begin{aligned} PV &= PV^{a;m+n} - PV^{a;m} = \bar{a}_{\overline{T_x \wedge (m+n)}|} - \bar{a}_{\overline{T_x \wedge m}|} \\ &= \frac{e^{-r(T_x \wedge m)} - e^{-r(T_x \wedge (m+n))}}{r} \end{aligned} \tag{4.21}$$

The expected present value is

$$_{m|n}\bar{a}_x = \bar{a}_{x\,\overline{m+n}|} - \bar{a}_{x\,\overline{m}|} = \int_m^{m+n} e^{-rt}\,{}_t p_x\, dt = {}_m E_x\, \bar{a}_{x+m\,\overline{n}|}\,. \tag{4.22}$$

The last expression can be obtained also by the rule of iterated expectation, and we carry through this small exercise just to illustrate the technique:

$$\begin{aligned} \mathbb{E}\,[PV] &= \mathbb{E}\,\mathbb{E}\,[PV \mid I_m] \\ &= {}_m p_x\, \mathbb{E}\,[PV \mid I_m = 1] + {}_m q_x\, \mathbb{E}\,[PV \mid I_m = 0] \\ &= {}_m p_x\, v^m\, \bar{a}_{x+m\,\overline{n}|}\,. \end{aligned}$$

An m-year deferred whole life annuity is obtained by putting $n = \infty$. The expected value is denoted by $_{m|}\bar{a}_x$.

Deferred life assurances, although less common in practice, are defined likewise. For instance, an m-year deferred n-year term assurance of 1 is payable upon death in the time interval $(m, m+n]$. Its present value at time 0 is

$$PV = PV^{ti;m+n} - PV^{ti;m}\,, \tag{4.23}$$

and its expected present value is

$$_{m|n}\bar{A}_{\overset{1}{x}} = \bar{A}_{\overset{1}{x}\,\overline{m+n|}} - \bar{A}_{\overset{1}{x}\,\overline{m|}} = {}_mE_x\,\bar{A}_{\overset{1}{x+m}\,\overline{n|}}\,,. = \int_m^{m+n} e^{-r\tau}\,{}_\tau p_x\,\mu_{x+\tau}\,d\tau\,. \tag{4.24}$$

F. Computational aspects. Distribution functions of present values and many other functions of interest can be calculated easily; after all there is only one random variable in play, and finding expected values amounts just to forming integrals in one dimension. We shall, however, not pursue this approach because it will turn out that a different point of view is needed in more complex situations to be studied in the sequel.

Table 4.1: Expected value (E), coefficient of variation (CV), and skewness (SK) of the present value at time 0 of a pure endowment (PE) with sum 1, a term insurance (TI) with sum 1, an endowment insurance (EI) with sum 1, and a life annuity (LA) with level intensity 1 per year, when $x = 30$, $n = 30$, μ is given by (3.25), and $r = \ln(1.045)$.

	PE	TI	EI	LA
E	0.2257	0.06834	0.2940	16.04
CV	0.4280	2.536	0.3140	0.1308
SK	−1.908	2.664	4.451	−4.451

Anyway, by methods to be developed later, we easily compute the three first moments of the present values considered above, and find their expected values, coefficients of variation, and skewnesses shown in Table 4.1. The reader should contemplate the results, keeping in mind that the coefficient of variation may be taken as a simple measure of "riskiness".

We interpose that numerical techniques will be dominant in our context. Explicit formulas cannot be obtained even for trivial quantities like $\bar{a}_{x\,\overline{n|}}$ under the Gompertz-Makeham law (3.23); age dependence and other forms of inhomogeneity of basic entities leave little room for aesthetics in actuarial science. Also relationships like (4.18) are of limited interest; they are certainly not needed for computational purposes, but may provide some general insight.

4.2 The principle of equivalence

A. A note on terminology. Like any other good or service, insurance coverage is bought at some price. And, like any other business, an insurance company must fix prices that are sufficient to defray the costs. In one respect, however, insurance is different: for obvious reasons the customer is to pay in advance. This circumstance is reflected by the insurance terminology, according to which payments made by the insured are called *premiums.* This word has the positive connotation "prize" (reward), rather antonymous to "price" (sacrifice, due), but the etymological background is, of course, that premium means "first" (French: prime).

B. The equivalence principle. The equivalence principle of insurance states that the expected present values of premiums and benefits should be equal. Then, roughly speaking, premiums and benefits will balance on the average. This idea will be made precise later. For the time being all calculations are made on an *individual net basis*, that is, the equivalence principle is applied to each individual policy, and without regard to expenses incurring in addition to the benefits specified by the insurance treaties. The resulting premiums are called (individual) *net premiums.*

The premium rate depends on the premium payment scheme. In the simplest case, the full premium is paid as a single amount immediately upon the inception of the policy. The resulting *net single premium* is just the expected present value of the benefits, which for basic forms of insurance is given in Section 4.1.

The net single premium may be a considerable amount and may easily exceed the liquid assets of the insured. Therefore, premiums are usually paid by a series of installments extending over some period of time. The most common solution is to let a fixed level amount fall due periodically, e.g. annually or monthly, from the inception of the agreement until a specified time m and contingent on the survival of the insured. Assume for the present that the premiums are paid continuously at a fixed level rate π. (This is admittedly an artificial assumption, but it can serve well as an approximate description of periodical payments, which will be treated later.) Then the premiums form an m-year temporary life annuity, payable by the insured to the insurer. Its present value is $\pi PV^{a;m}$, with expected value $\pi\bar{a}_{x\,\overline{m}|}$ given by (4.14). We list formulas for the net level premium rate for a selection of basic forms of insurance:

For the pure endowment (Paragraph 4.1.B) against level premium in the insurance period,

$$\pi = \frac{{}_nE_x}{\bar{a}_{x\,\overline{n}|}}\,. \tag{4.25}$$

For the term insurance (Paragraph 4.1.C) against level premium in the insurance period,

$$\pi = \frac{\bar{A}^{\,1}_{x\,\overline{n}|}}{\bar{a}_{x\,\overline{n}|}}\,. \tag{4.26}$$

For the endowment insurance (Paragraph 4.1.C) against level premium in the insurance period,

$$\pi = \frac{\bar{A}_{x\,\overline{n}|}}{\bar{a}_{x\,\overline{m}|}} = \frac{1}{\bar{a}_{x\,\overline{n}|}} - r\,, \tag{4.27}$$

the last expression following from (4.18).

For the m-year deferred n-year temporary life annuity (Paragraph 4.1.E) against level premium in the deferred period,

$$\pi = \frac{{}_{m|n}\bar{a}_x}{\bar{a}_{x\,\overline{m}|}} = \frac{\bar{a}_{x\,\overline{m+n}|}}{\bar{a}_{x\,\overline{m}|}} - 1\,. \tag{4.28}$$

C. The net economic result for a policy. The random variables studied in Section 4.1 represent the uncertain future liabilities of the insurer. Now, unless single premiums are used, also the premium incomes are dependent on the insured's life length and become a part of the insurer's uncertainty. Therefore, the relevant random variable associated with an insurance policy is the present value of benefits less premiums,

$$PV = PV^b - \pi PV^{a;m}\,, \tag{4.29}$$

where PV^b is the present value of the benefits, e.g. $PV^{ei;n}$ in the case of an n-year endowment insurance.

Stated precisely, the equivalence principle lays down that

$$\mathbb{E}[PV] = 0\,. \tag{4.30}$$

For example, with $PV^b = PV^{ei;n}$ (4.30) becomes $0 = \bar{A}_{x\,\overline{n}|} - \pi\bar{a}_{x\overline{m}|}$, which yields (4.27) when $m = n$.

A measure of the uncertainty associated with the economic result of the policy is the variance $\mathbb{V}[PV]$. For example, with $PV^b = PV^{ei;n}$ and $m = n$,

$$\begin{aligned}
\mathbb{V}[PV] &= \mathbb{V}\left[v^{T_x \wedge n} - \pi\,\frac{1 - v^{T_x \wedge n}}{r}\right] = (1 + \pi/r)^2\mathbb{V}[v^{T_x \wedge n}] \\
&= \frac{2\left(\bar{a}_{x\,\overline{n}|} - \bar{a}^{(2r)}_{x\,\overline{n}|}\right)}{r\bar{a}^2_{x\,\overline{n}|}} - 1\,.
\end{aligned} \tag{4.31}$$

4.3 Prospective reserves

A. The case. We shall discuss the notion of reserve in the framework of a combined insurance which comprises all standard forms of contingent payments that have been studied so far and, therefore, easily specializes to any of those. The insured is x years old upon issue of the contract, which is for a term of n years. The benefits consist of a term insurance with sum insured b_t payable upon death at time $t \in (0, n)$ and a pure endowment with sum b_n payable upon survival at time n. The premiums consist of a lump sum π_0 payable immediately upon the inception of the policy at time 0, and thereafter an annuity payable continuously at rate π_t per time unit contingent on survival at time $t \in (0, n)$. As before, assume that the interest rate is a deterministic function r_t.

The expected present value at time 0 of total benefits less premiums under the contract can be put up directly as the sum of the expected discounted payments in each small time interval:

$$-\pi_0 + \int_0^n e^{-\int_0^\tau r}\, {}_\tau p_x \{\mu_{x+\tau}\, b_\tau - \pi_\tau\}\, d\tau + b_n\, e^{-\int_0^n r}\, {}_n p_x\,. \tag{4.32}$$

Under the equivalence principle this is set equal to 0, a constraint on the premium function π.

B. Definition of the reserve. The expected value (4.32) represents, in an average sense, an assessment of the economic prospects of the policy at the outset. At any time $t > 0$ in the subsequent development of the policy the assessment should be updated with regard to the information currently available. If the policy has expired by death before time t, there is nothing more to be done. If the policy is still in force, a renewed assessment must be based on the conditional distribution of the remaining life length. Insurance legislation lays down that at any time the insurance company must provide a reserve to meet future net liabilities on the contract, and this reserve should be precisely the expected present value at time t of total benefits less premiums in the future. Thus, if the policy is still in force at time t, the reserve is

$$V_t = \int_t^n e^{-\int_t^\tau r}\, {}_{\tau-t} p_{x+t} \{\mu_{x+\tau}\, b_\tau - \pi_\tau\}\, d\tau + b_n\, e^{-\int_t^n r}\, {}_{n-t} p_{x+t}\,. \tag{4.33}$$

More precisely, this quantity is called the *prospective reserve* at time t since it "looks ahead". Under the principle of equivalence it is usually called *the net premium reserve.* We will take the liberty to just speak of the *reserve.*

Upon inserting ${}_{\tau-t}p_{x+t} = e^{-\int_t^\tau \mu_{x+s}\, ds}$, (4.33) assumes the form

$$V_t = \int_t^n e^{-\int_t^\tau (r_s+\mu_{x+s})\, ds} \{\mu_{x+\tau}\, b_\tau - \pi_\tau\}\, d\tau + e^{-\int_t^\tau (r_s+\mu_{x+s})\, ds} b_n\,. \tag{4.34}$$

Glancing behind at (2.14), we see that, formally, the expression in (4.34) is the reserve at time t for a deterministic contract with payments given by $\Delta B_0 = -\pi_0$ (comes from setting (4.32) equal to 0), $dB_t = (\mu_{x+t}\, b_t - \pi_t)\, dt$, $0 < t < n$ and $\Delta B_n = b_n$, and with interest rate $r_t + \mu_{x+t}$. We can, therefore, reuse the relationships in Chapter 2.

For instance, by (2.26) and (2.13), we have the following *retrospective formula* for the prospective premium reserve:

$$V_t = e^{\int_0^t (r_s+\mu_{x+s})\, ds}\pi_0 + \int_0^t e^{\int_\tau^t (r_s+\mu_{x+s})\, ds}(\pi_\tau - \mu_{x+\tau}\, b_\tau)\, d\tau\,. \tag{4.35}$$

This formula expresses V_t as the surplus of transactions in the past, accumulated at time t with the "benefit of interest and survivorship".

C. Some special cases. The net reserve is easily put up for the various forms of insurance treated in Sections 4.1 and 4.2. We assume that the interest rate is

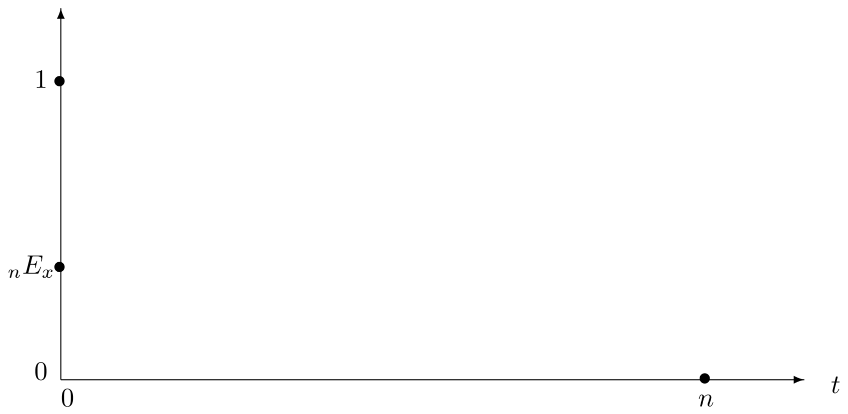


Figure 4.1: The net reserve for an n-year pure endowment of 1 against single net premium.

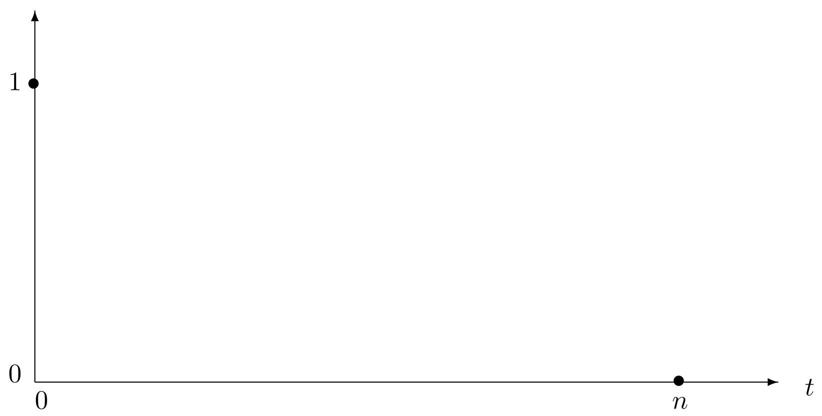


Figure 4.2: The net reserve for an n-year pure endowment of 1 against level premium in the insurance period.

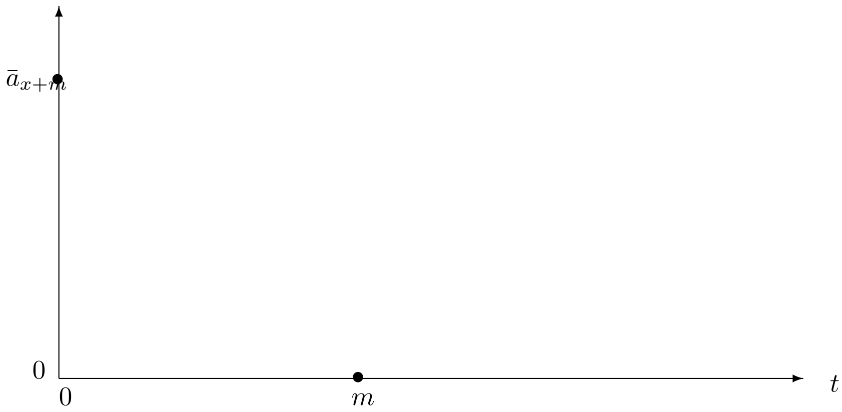


Figure 4.3: The net reserve for an m-year deferred whole life annuity against level premium in the deferred period.

constant and that premiums are based on the equivalence principle, which can be expressed as

$$V_0 = \pi_0 \,. \tag{4.36}$$

First, for the pure endowment against single net premium ${}_nE_x$ collected at time 0,

$$V_t = {}_{n-t}E_{x+t}, \;\; 0 \le t < n. \tag{4.37}$$

The graph of V_t will typically look as in Fig. 4.1. At points of discontinuity a dot marks the value of the function.

If premiums are payable continuously at level rate π given by (4.25) throughout the insurance period, then

$$\begin{aligned} V_t &= {}_{n-t}E_{x+t} - \pi \bar{a}_{x+t\,\overline{n-t|}} \\ &= {}_{n-t}E_{x+t} - \frac{{}_nE_x}{\bar{a}_{x\,\overline{n|}}}\bar{a}_{x+t\,\overline{n-t|}}. \end{aligned} \tag{4.38}$$

A typical graph of this function is shown in Fig. 4.2.

Next, for an m-year deferred whole life annuity against level net premium π given by (4.28),

$$\begin{aligned} V_t &= \begin{cases} {}_{m-t|}\bar{a}_{x+t} - \pi \bar{a}_{x+t\,\overline{m-t|}}, & 0 < t < m, \\ \bar{a}_{x+t}, & t \ge m, \end{cases} \\ &= \bar{a}_{x+t} - \bar{a}_{x+t\,\overline{m-t|}} - \frac{\bar{a}_x - \bar{a}_{x\,\overline{m|}}}{\bar{a}_{x\,\overline{m|}}}\bar{a}_{x+t\,\overline{m-t|}} \\ &= \bar{a}_{x+t} - \frac{\bar{a}_x}{\bar{a}_{x\,\overline{m|}}}\bar{a}_{x+t\,\overline{m-t|}} \end{aligned} \tag{4.39}$$

(with the understanding that $\bar{a}_{x\,\overline{m-t|}} = 0$ if $t > m$). A typical graph of this function is shown in Fig. 4.3.

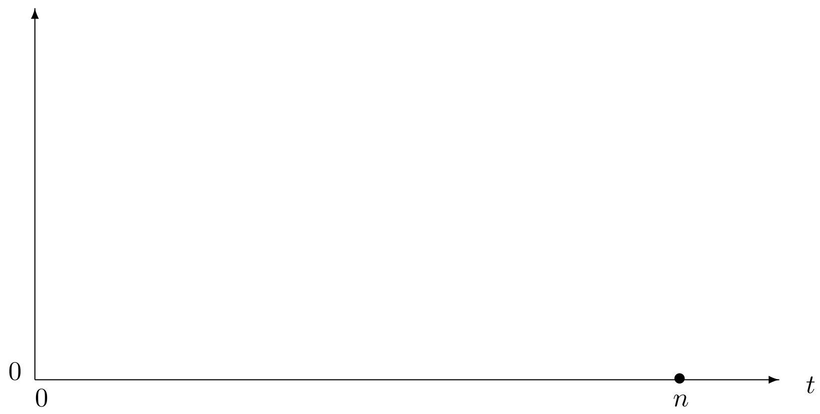


Figure 4.4: The net reserve for an n-year term insurance against level premium in the insurance period

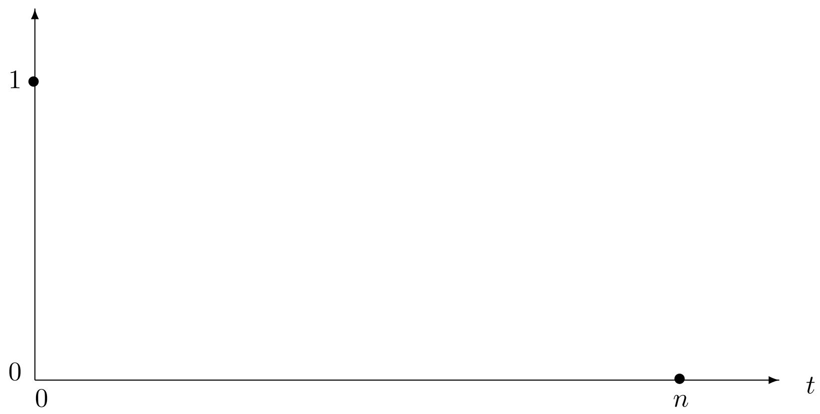


Figure 4.5: The net reserve for an n-year endowment insurance with level premium payable in the insurance period.

For the n-year term insurance against level net premium π given by (4.26),

$$\begin{aligned} V_t &= \bar{A}_{\overset{1}{x+t}\,\overline{n-t}|} - \pi \bar{a}_{x+t\,\overline{n-t}|} \\ &= 1 - r\bar{a}_{x+t\,\overline{n-t}|} - {}_{n-t}E_{x+t} - \frac{1 - r\bar{a}_{x\,\overline{n}|} - {}_nE_x}{\bar{a}_{x\,\overline{n}|}} \bar{a}_{x+t\,\overline{n-t}|} \\ &= 1 - {}_{n-t}E_{x+t} - (1 - {}_nE_x)\frac{\bar{a}_{x+t\,\overline{n-t}|}}{\bar{a}_{x\,\overline{n}|}}. \end{aligned} \tag{4.40}$$

A typical graph of this function is shown in Fig. 4.4.

Finally, for the n-year endowment insurance against level net premium π given by (4.27),

$$\begin{aligned} V_t &= \bar{A}_{x+t\,\overline{n-t}|} - \pi \bar{a}_{x+t\,\overline{n-t}|} \\ &= 1 - r\bar{a}_{x+t\,\overline{n-t}|} - \frac{1 - r\bar{a}_{x\,\overline{n}|}}{\bar{a}_{x\,\overline{n}|}} \bar{a}_{x+t\,\overline{n-t}|} \\ &= 1 - \frac{\bar{a}_{x+t\,\overline{n-t}|}}{\bar{a}_{x\,\overline{n}|}}. \end{aligned} \tag{4.41}$$

A typical graph of this function is shown in Fig. 4.5.

The reserve in (4.41) is, of course, the sum of the reserves in (4.39) and (4.40). Note that the pure term insurance requires a much smaller reserve than the other insurance forms, with elements of savings in them. However, at old ages x (where people typically are not covered against the risk of death since death will incur soon with certainty) also the term insurance may have a V_t close to 1 in the middle of the insurance period.

D. Non-negativity of the reserve. In all the examples given here the net reserve is sketched as a non-negative function. Non-negativity of V_t is not a consequence of the definition. One may easily construct premium payment schemes that lead to negative values of V_t (just let the premiums fall due after the payment of the benefits), but such payment schemes are not used in practice. The reason is that the holder of a policy with $V_t < 0$ is in expected debt to the insurer and would thus have an incentive to cancel the policy and thereby get rid of the debt. (The agreement obliges the policy-holder only to pay the premiums, and the contract can be terminated at any time the policy-holder wishes.) Therefore, it is in practice required that

$$V_t \geq 0, \quad t \geq 0. \tag{4.42}$$

E. The reserve considered as a function of time. We will now take a closer look at the prospective reserve as a function of time, bearing in mind that it should be non-negative. The building blocks are the expected present values ${}_{n-t}E_{x+t}$, $\bar{a}_{x+t:\overline{n-t}|}$, $\bar{A}_{x+t\,\overline{n-t}|}$, and $\bar{A}_{\overset{1}{x+t}\,\overline{n-t}|}$ appearing in the formulas in Section 4.3.

First,

$${}_{n-t}E_{x+t} = e^{-\int_t^n (r+\mu_{x+s})\,ds}$$

is seen to be an increasing function of t no matter what are the interest rate and the mortality rate. The derivative is

$$\frac{d}{dt}{}_{n-t}E_{x+t} = {}_{n-t}E_{x+t}\,(r + \mu_{x+t})\,.$$

We interpose here that nothing is changed if r depends on time. The expressions above show that, for this pure survival benefit, r and μ play identical parts in the expected present value. Thus, mortality bequest acts as an increase of the interest rate.

Next consider

$$\bar{a}_{x+t:\overline{n-t|}} = \int_t^n e^{-\int_t^\tau (r+\mu_{x+s})\,ds}\,d\tau\,.$$

The following inequalities are obvious:

$$\bar{a}_{x+t:\overline{n-t|}} \leq \frac{1}{r + \inf_{s\geq t}\mu_{x+s}} \leq \frac{1}{r}\,.$$

The last expression is just the present value of a perpetuity, (2.32). If μ is an increasing function, then

$$\bar{a}_{x+t:\overline{n-t|}} \leq \frac{1}{r + \mu_{x+t}}\,.$$

We find the derivative

$$\frac{d}{dt}\bar{a}_{x+t:\overline{n-t|}} = (r + \mu_{x+t})\bar{a}_{x+t:\overline{n-t|}} - 1\,.$$

It follows that $\bar{a}_{x+t:\overline{n-t|}}$ is a decreasing function of t if μ is increasing, which is quite natural. You can easily invent an example where $\bar{a}_{x+t:\overline{n-t|}}$ is not decreasing.

From the identity

$$\bar{A}_{x+t\,\overline{n-t|}} = 1 - r\bar{a}_{x+t:\overline{n-t|}}$$

we conclude that $\bar{A}_{x+t\,\overline{n-t|}}$ is an increasing function of t if μ is increasing.

For

$$\bar{A}_{\overset{1}{x+t}\,\overline{n-t|}} = 1 - r\bar{a}_{x+t:\overline{n-t|}} - {}_{n-t}E_{x+t}$$

no general statement can be made as to whether it is decreasing or increasing.

Looking back at the formulas derived in Paragraph C above, we can conclude that the reserve for the pure life endowment against single premium, (4.37), is always increasing. Assume henceforth that μ is increasing, as is usually the case at ages when people are insured and certainly holds for the Gompertz-Makeham law. Then also the reserve (4.38) for the pure life endowment against level premium during the term of the contract is increasing, and the same is the case for the reserve (4.41) of the endowment insurance. It is left to the diligent reader to show that the reserve in (4.39) is increasing throughout the deferred period and thereafter turns decreasing. This is best done by examining (4.33) and (4.35) in the cases $t > m$ and $t \leq m$, respectively. It follows in particular that the reserve is non-negative. The same trick serves also to show that (4.40) is first increasing and thereafter decreasing.

4.4 Thiele's differential equation

A. The differential equation. We turn back to the general case with the reserve given by (4.33) or (4.34), the latter being the more convenient since we can draw on the results in Chapter 2.

The differential form (2.19) translates to the celebrated *Thiele's differential equation*,

$$\frac{d}{dt}\,V_t = (r + \mu_{x+t})\,V_t + \pi_t - \mu_{x+t}\,b_t, \tag{4.43}$$

valid at each t where b, π, and μ are continuous. The right hand side expression in (4.43) shows how the fund per surviving policy-holder changes per time unit at time t. It is increased by the interest earned, rV_t, by the fund inherited from those who die, $\mu_{x+t}\,V_t$, and by the excess of premiums over expected benefits (which may be negative, of course).

When combined with the boundary condition

$$V_{n-} \quad = \quad b_n\,, \tag{4.44}$$

the differential equation (4.43) determines V_t for fixed b and π.

If the principle of equivalence is exercised, then we must add the condition (4.36). This represents a constraint on the contractual payments b and π; typically, one first specifies the benefit b and then determines the premium rate for a given premium plan (shape of π).

Thiele's differential equation is a so-called *backward differential equation.* This term indicates that we take our stand at the beginning of the time interval we are interested in and also that the differential equation is to be solved by a backward scheme starting from the ultimate condition (2.21). The differential equation may be put up by the *direct backward construction* which goes as follows. Suppose the policy is in force at time $t \in (0, n)$. Use the rule of iterated expectation, conditioning on what happens in the small time interval $(t, t + dt]$: with probability $\mu_{x+t}\,dt + o(dt)$ the insured dies, and the conditional expected value is then just b_t; with probability $1 - \mu_{x+t}\,dt + o(dt)$ the insured survives, and the conditional expected value is then $-\pi_t dt + e^{-r\,dt}\,V_{t+dt}$. We gather

$$V_t = b_t\,\mu_{x+t}\,dt - \pi_t\,dt + (1 - \mu_{x+t}\,dt)e^{-r\,dt}\,V_{t+dt} + o(dt). \tag{4.45}$$

Subtract V_{t+dt} on both sides, divide by dt and let dt tend to 0. Observing that $(e^{-rdt} - 1)/dt \to -r$ as $dt \to 0$, one arrives at (4.43)

B. Savings premium and risk premium. Suppose the equivalence principle is in use. Rearrange (4.43) as

$$\pi_t = \frac{d}{dt}\,V_t - rV_t + (b - V_t)\mu_{x+t}. \tag{4.46}$$

This form of the differential equation shows how the premium at any time decomposes into a *savings premium*,

$$\pi_t^s = \frac{d}{dt}\,V_t - rV_t, \tag{4.47}$$

and a *risk premium*,

$$\pi_t^r = (b_t - V_t)\mu_{x+t}. \tag{4.48}$$

The savings premium provides the amount needed in excess of the earned interest to maintain the reserve. The risk premium provides the amount needed in excess of the available reserve to cover an insurance claim.

C. Uses of the differential equation. In the examples given above, Thiele's differential equation was useful primarily as a means of investigating the development of the reserve. It was not required in the construction of the premium and the reserve, which could be put up by direct prospective reasoning. In the final example to be given Thiele's differential equation is needed as a constructive tool.

Assume that the pension treaty studied above is modified so that the reserve is paid back at the moment of death in case the insured dies during the contract period, the philosophy being that "the savings belong to the insured". Then the scheme is supplied by an $(n+m)$-year temporary term insurance with sum $b_t = V_t$ at any time $t \in (0, m+n)$. The solution to (4.43) is easily obtained as

$$V_t = \begin{cases} \pi \bar{s}_{\overline{t}|}, & 0 < t < m, \\ b\bar{a}_{\overline{m+n-t}|}, & m < t < m+n, \end{cases}$$

where $\bar{s}_{\overline{t}|} = \int_0^t (1+i)^{t-\tau}\, d\tau$. The reserve develops just as for ordinary savings contracts offered by banks.

D. Dependence of the reserve on the contract elements. A small collection of results due to Lidstone (1905) and, in the continuous time set-up, Norberg (1985), deal with the dependence of the reserve on the contract elements, in particular mortality and interest.

The starting point in the continuous time case is Thiele's differential equation. For the sake of concreteness, we adopt the model assumptions and the contract described in Section 4.4 and will refer to this as the *standard contract*. For ease of reference we fetch Thiele's differential from (4.43):

$$\frac{d}{dt}\, V_t = \pi_t - \mu_{x+t}\, b_t + (r_t + \mu_{x+t})\, V_t\, . \tag{4.49}$$

The boundary condition following from the very definition of the reserve is

$$V_{n-} = b_n\, . \tag{4.50}$$

With premiums determined by the principle of equivalence, we also have

$$V_0 = \pi_0\, , \tag{4.51}$$

where π_0 is the lump sum premium payment collected upon the inception of the policy (it may be 0, of course).

Now consider a different model with interest r_t^* and mortality μ_{x+t}^* and a different contract with benefits b_t^* and premiums π_t^*. This will be referred to as the *special contract*. The reserve function V_t^* under this contract satisfies

$$\frac{d}{dt} V_t^* = \pi_t^* - \mu_{x+t}^* b_t^* + (r_t^* + \mu_{x+t}^*) V_t^* , \tag{4.52}$$

$$V_{n-}^* = b_n^* , \tag{4.53}$$

$$V_0^* = \pi_0^* . \tag{4.54}$$

Assume that

$$\pi_0^* = \pi_0 , \qquad b_n^* = b_n . \tag{4.55}$$

We are interested in the difference $V_t^* - V_t$, and a few words are in order to motivate this: The reserve is accounted as a liability on the part of the insurance company. To be on the safe side, the company should, at any time, provide a reserve in excess of what seems likely to be needed. This is usually obtained by using 'technical' elements r_t^* and μ_{x+t}^* that are different from the 'realistic' elements r_t and μ_{x+t}, and that produce a reserve V_t^* bigger than the 'realistic' V_t.

Subtract (4.49) from (4.52) to get

$$\frac{d}{dt} (V_t^* - V_t) = \eta_t + (r_t^* + \mu_{x+t}^*) (V_t^* - V_t) , \tag{4.56}$$

where

$$\eta_t = (\pi_t^* - \pi_t) + (\mu_{x+t} b_t - \mu_{x+t}^* b_t^*) + (r_t^* - r_t + \mu_{x+t}^* - \mu_{x+t}) V_t . \tag{4.57}$$

Integrate (4.56) from 0 to t, using $V_0 = V_0^*$, to obtain

$$V_t^* - V_t = \int_0^t e^{\int_s^t (r^* + \mu^*)} \eta_s \, ds .$$

Similarly, integrate from t to n, using $V_{n-} = V_{n-}^*$, to obtain

$$V_t^* - V_t = - \int_t^n e^{-\int_t^s (r^* + \mu^*)} \eta_s \, ds .$$

From these relations conclude: If there exists a $t_0 \in [0, n]$ such that

$$\eta_t \ \begin{matrix} \leq \\ \geq \end{matrix} \ 0 \ \text{ for } \ t \ \begin{matrix} < \\ > \end{matrix} \ t_0 , \tag{4.58}$$

then $V_t^* \leq V_t$ for all t. In particular, this is the case if η_t is non-decreasing. The result remains valid if all inequalities are made strict. We can now prove the following:

(1) For a contract with level premium intensity throughout the insurance period, and with non-decreasing reserve, a uniform increase of the interest rate results in a decrease of the reserve.

Proof: Now $r_t^* - r_t = \Delta r$ is a positive constant, π^* and π are both constants, all other elements are unchanged, and V_t increasing. Then $\eta_t = (\pi^* - \pi) + \Delta r\, V_t$ is increasing.

(2) Consider an endowment insurance with fixed sum insured and level premium rate throughout the insurance period. Prove that a change of mortality from μ to μ^* such that $\mu_t^* - \mu_t$ is positive and non-increasing, leads to a decrease of the reserve.

Proof: Now $(\mu_{x+t}^* - \mu_{x+t})$ is positive and decreasing (non-increasing), π^* and π are constants, $b_t = b_t^* = b$ constant, V_t increasing (this is the case for the endowment insurance if μ in increasing). Then, since $V_t \leq b$, we have $\eta_t = (\pi^* - \pi) - (\mu_{x+t}^* - \mu_{x+t})(b - V_t)$ is increasing.

(3) Consider a policy with no down premium payment at time 0 and no life endowment at time n. Let the special contract be the same as the standard one, except that the special contract charges so-called natural premium, $\pi_t^* = b_t \mu_{x+t}$. Then $V_t^* = 0$ for all t, and (4.58) can be used to check whether the reserve V_t is non-negative (as it should be).

Proof: Putting $\pi_t^* = \mu_{x+t} b_t$, means premiums covers current expected benefits, so there is no accumulation of reserve; $V_t^* = 0$. Now $\eta_t = -\pi_t + \mu_{x+t} b_t$, so if this is increasing, then $0 = V_t^* \leq V_t$. This is the case e.g. if π and b are constants and μ_t is increasing.

The reason why the impact on the reserve of a change in valuation and/or contract elements is a bit involved is that, under the equivalence principle, the premium is also affected by the change. However, if we require that the premium be constant as function of t, then $(\pi^* - \pi)$ appearing in the expression for η_t is constant and does not affect the monotonicity properties of η_t. Note also that, since $V_t^* - V_t$ starts and ends at 0, η_t cannot be strictly positive in some part of $(0, n)$ without being strictly negative in some other part.

4.5 Probability distributions

A. Motivation. The basic paradigm being the principle of equivalence, life insurance mathematics centers on expected present values. The key tool is Thiele's differential equation, which describes the development of such expected values and forms a basis for computing them by recursive methods. In Chapter 7 we shall obtain analogous differential equations for higher order moments, which will enable us to compute the variance, skewness, kurtosis, and so on of the present value of payments under a fairly general insurance contract.

We shall give an example of how to determine the probability distribution of a present value, which is at the base of the moments and of any other expected values of interest. Knowledge of this distribution, and in particular its upper tail, gives insight into the riskiness of the contract beyond what is provided by the mean and some higher order moments.

The task is easy for an insurance on a single life since then the model involves only one random variable (the life length of the insured). De Pril [14] and Dhaene [16] offer a number of examples. In principle the task is simple also for insurances involving more than one life or, more generally, a finite number of random variables. In such situations the distributions of present values (and any other functions of the random variables) can be obtained by integrating the finite-dimensional distribution.

B. A simple example. Consider the single life status (x) with remaining life time T_x distributed as described in Chapter 3. Suppose (x) buys an n year term insurance with fixed sum b and premiums payable continuously at level rate π per year as long as the contract is in force (see Paragraphs 4.1.C-D). The present value of benefits less premiums on the contract is

$$U(T_x) = be^{-rT_x} 1[0 < T_x < n] - \pi \bar{a}_{\overline{T_x \wedge n}|} \,,$$

where $\bar{a}_{\overline{t}|} = \int_0^t e^{-r\tau}\, d\tau = (1 - e^{-rt})/r$ is the present value of an annuity certain payable continuously at level rate 1 per year for t years. The function U is non-increasing in T_x, and we easily find the probability distribution

$$\mathbb{P}[U \le u] = \begin{cases} 0 & , \quad u < -\pi \bar{a}_{\overline{n}|}, \\ \mathbb{P}[T_x > n] & , \quad -\pi \bar{a}_{\overline{n}|} \le u < b\, e^{-rn} - \pi \bar{a}_{\overline{n}|}, \\ \mathbb{P}\left[T_x > \frac{1}{r} \ln \left(\frac{br+\pi}{ur+\pi}\right)\right] & , \quad be^{-rn} - \pi \bar{a}_{\overline{n}|} \le u < b, \\ 1 & , \quad u \ge b. \end{cases} \tag{4.59}$$

The jump at $-\pi \bar{a}_{\overline{n}|}$ is due to the positive probability of survival to time n. Similar effects are to be anticipated also for other insurance products with a finite insurance period since, in general, there is a positive probability that the policy will remain in the current state until the contract terminates.

The probability distribution in (4.59) is depicted in Fig. 4.6 for the G82M case with $r = \ln(1.045)$ and $\mu(t|x) = 0.0005 + 10^{-4.12+0.038(x+t)}$ when $x = 30$, $n = 30$, $b = 1$, and $\pi = 0.0042608$ (the equivalence premium).

4.6 The stochastic process point of view

A. The processes indicating survival and death. In Paragraph A of Section 4.1 we introduced the indicator of the event of survival to time t, $I_t = 1[T_x > t]$, and the indicator of the complementary event of death within time t, $N_t = 1 - I_t = 1[T_x \le t]$. Viewed as functions of t, they are stochastic processes. The latter counts the number of deaths of the insured as time progresses and is thus a simple example of a counting process as defined in Paragraph D of Appendix A. This motivates the notation N_t. By their very definitions, I_t and N_t are RC.

In the present context, where everything is governed by just one single random variable, T_x, the process point of view is not important for practical purposes. For didactic purposes, however, it is worthwhile taking it already here as

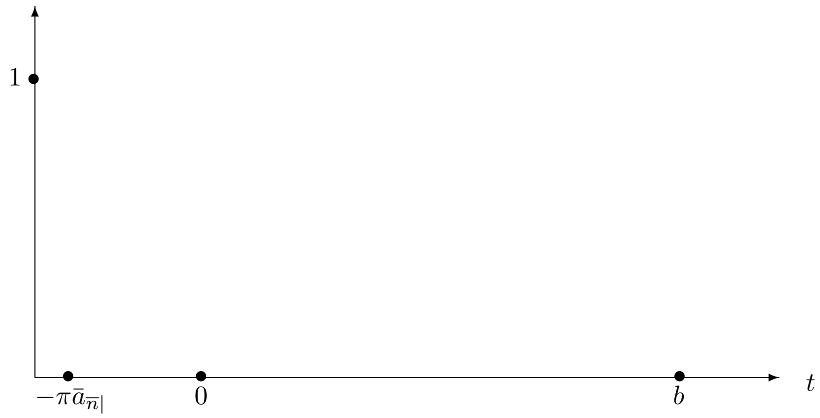


Figure 4.6: The probability distribution of the present value of a term insurance against level premium.

a rehearsal for more complicated situations where stochastic processes cannot be dispensed with.

The payment functions of the benefits considered in Section 4.1 can be recast in terms of the processes I_t and N_t. In differential form they are

$$\begin{aligned} dB_t^{e;n} &= I_t\, d\varepsilon_n(t)\,, \\ dB_t^{ti;n} &= 1_{(0,n]}(t)\, dN_t\,, \\ dB_t^{a;n} &= I_t 1_{(0,n)}(t)\, dt\,, \\ dB_t^{ei;n} &= dB_t^{ti;n} + dB_t^{e;n}\,. \end{aligned}$$

Their present values at time 0 are

$$\begin{aligned} PV^{e;n} &= e^{-\int_0^n r} I_n\,, \\ PV^{ti;n} &= \int_0^n e^{-\int_0^\tau r} dN_\tau\,, \\ PV^{a;n} &= \int_0^n e^{-\int_0^\tau r} I_\tau d\tau\,, \\ PV^{ei;n} &= V^{ti;n} + V^{e;n}\,. \end{aligned}$$

The expressions in (4.14) and (4.10) are obtained directly by taking expectation under the integral sign, using the obvious relations

$$\begin{aligned} \mathbb{E}\left[I_\tau\right] &= {}_\tau p_x\,, \\ \mathbb{E}\left[dN_\tau\right] &= {}_\tau p_x\, \mu_{x+\tau}\, d\tau\,. \end{aligned}$$

The relationship (4.19) re-emerges in its more basic form upon integrating by parts to obtain

$$e^{-\int_0^n r} I_n = 1 + \int_0^n e^{-\int_0^\tau r}(-r_\tau) I_\tau\, d\tau + \int_0^n e^{-\int_0^\tau r} dI_\tau\,,$$

and setting $dI_t = -dN_t$ in the last integral.

Chapter 5

Expenses

5.1 A single life insurance policy

A. Three categories of expenses. Every business must defray expenditures that come in addition to the net production costs of the commodities or services it offers, and these expenses must be reckoned in the prices paid by the customers. Thus, the premiums charged for a given insurance contract must cover, not only the contractual net benefits, but also all items of expenditure connected with the operations of the insurance company.

For the sake of concreteness, and also of loyalty to standard actuarial notation, we shall introduce the issue of expenses in the framework of the single life policy treated in Chapter 4. To create a case that involves all types of payments, let us consider a life (x) who purchases an n-year endowment insurance with sum insured b and premium payable continuously at level rate as long as the policy is in force.

From Chapter 4 we retrieve the following relationships for the premium and the reserve calculated net of administration expenses, henceforth termed *net premium* and *net premium reserve*. Upon scaling with the sum insured b, formula (4.27) for the net premium rate becomes

$$\pi = b\frac{\bar{A}_{x\,\overline{n}|}}{\bar{a}_{x\,\overline{n}|}} = b\left(\frac{1}{\bar{a}_{x\,\overline{n}|}} - r\right)\,, \tag{5.1}$$

and formula (4.41) for the net premium reserve becomes

$$V_t = b\bar{A}_{x+t\,\overline{n-t}|} - \pi\bar{a}_{x+t\,\overline{n-t}|} = b\left(1 - \frac{\bar{a}_{x+t\,\overline{n-t}|}}{\bar{a}_{x\,\overline{n}|}}\right). \tag{5.2}$$

From (4.43) and (4.44) we recall Thiele's differential equation

$$\frac{d}{dt}\,V_t = (r + \mu_{x+t})\,V_t + \pi - \mu_{x+t}b\,, \tag{5.3}$$

and its natural condition

$$V_{n-} \quad = \quad b\,. \tag{5.4}$$

The equivalence requirement (4.36) is now just

$$V_0 = 0\,. \tag{5.5}$$

When expenses are included in the accounts one will have to charge a *gross premium* rate π', which obviously must be greater than the net premium rate π. The *gross premium reserve* V_t' will also in general differ from its net counterpart V_t, but it is not obvious at the outset which of the two is greater since the gross reserve includes both greater expenses and greater premiums. The gross quantities can be defined precisely only after we have made specific assumptions about the structure of the expenses, which we now will do.

The expenses are usually divided into three categories. In the first place there are the so-called α-expenses that incur in connection with the establishment of the contract. They comprise acquisition costs including marketing and commission to the sales agent, and costs related to health examination, issue of the policy, entering the details of the contract into the data files, etc. It is assumed that these expenses incur immediately at time 0 and that they are of the form

$$\alpha' + \alpha'' b\,. \tag{5.6}$$

In the second place there are the so-called β-expenses that incur in connection with collection and accounting of premiums. They are assumed to incur continuously at constant rate

$$\beta' + \beta''\pi' \tag{5.7}$$

throughout the premium-paying period.

Finally, there are the so-called γ-expenses that comprise all expenditures not included in the former two categories, such as wages to employees, rent, taxes, fees, and maintenance of the business operations in general. These expenses are assumed to incur continuously at rate

$$\gamma' + \gamma'' b + \gamma''' V_t' \tag{5.8}$$

at time t if the policy is then in force.

The terms α', β', and γ' represent costs that are the same for all policies. The terms $\alpha'' b$, $\beta''\pi'$, and $\gamma'' b$ represent costs that are proportional to the size of the contract as measured by the amounts specified in the policy. Typically this is the case for the agent's commission, which may be a considerable portion of the α-expenses on individual insurances sold in an open competitive market, and also for the debt collector's or solicitor's commission, which in former days made up the major part of the β-expenses. The term $\gamma''' V_t'$ represents expenses in connection with management of the investment portfolio, which can reasonably be divided between the policy-holders in proportion to their current reserves.

B. The gross premium and the gross premium reserve. Upon exercising the equivalence principle in the presence of expenses, one will determine the *gross premium* rate π' and the corresponding *gross premium reserve* function V'_t. It is convenient to employ the Thiele technique because expenses depend on the reserve as specified in (5.8). Upon replacing the annuity type of premiums less benefits in (5.3) (which are just π) with $\pi' - \beta' - \beta''\pi' - \gamma' - \gamma''b - \gamma'''V'_t$, we obtain

$$\frac{d}{dt}V'_t = (r + \mu_{x+t})V'_t + \pi' - \beta' - \beta''\pi' - \gamma' - \gamma''b - \gamma'''V'_t - \mu_{x+t}b\,. \tag{5.9}$$

The natural side condition is

$$V'_{n-} = b\,, \tag{5.10}$$

and the equivalence requirement now becomes

$$V'_0 = -(\alpha' + \alpha''b)\,. \tag{5.11}$$

Gathering terms involving V'_t on the left of (5.9) and multiplying on both sides with $e^{\int_t^n (r-\gamma'''+\mu)}$ gives

$$\frac{d}{dt}\left(e^{\int_t^n (r-\gamma'''+\mu)}V'_t\right) = e^{\int_t^n (r-\gamma'''+\mu)}\{(1-\beta'')\pi' - \beta' - \gamma' - (\gamma'' + \mu_{x+t})b\}\,. \tag{5.12}$$

Now integrate (5.12) between t and n, using (5.10), and rearrange a bit to obtain

$$\begin{aligned} V'_t &= \int_t^n e^{-\int_t^\tau (r-\gamma'''+\mu)}\{\beta' + \gamma' + (\gamma'' + \mu_{x+\tau})b - (1-\beta'')\pi'\}\,d\tau \\ &\quad + e^{-\int_t^n (r-\gamma'''+\mu)}b\,. \end{aligned} \tag{5.13}$$

Upon inserting $t = 0$ into (5.13) and using (5.11), we find

$$\pi' = \frac{\alpha' + \alpha''b + \int_0^n e^{-\int_0^\tau (r-\gamma'''+\mu)}\{\beta' + \gamma' + (\gamma'' + \mu_{x+\tau})b\}d\tau + e^{-\int_0^n (r-\gamma'''+\mu)}b}{(1-\beta'')\int_0^n e^{-\int_0^\tau (r-\gamma'''+\mu)}d\tau}\,. \tag{5.14}$$

In the special case where $\gamma''' = 0$ we could determine π' and V'_t directly from the defining relations without using the differential equation. That goes, in fact, also for the general case with $\gamma''' \neq 0$ by the following consideration: By inspection of the differential equation (5.9) and the side conditions, it is realized that, formally, the problem amounts to determining the "net premium rate" $(1-\beta'')\pi'$ and "net premium reserve" V'_t for a policy with (admittedly unrealistic) benefits consisting of a lump sum payment of $\alpha' + \alpha''b$ at time 0, a continuous level life annuity of $\beta' + \gamma' + \gamma''b$ per year, and an endowment insurance of b, when the interest rate is $r - \gamma'''$.

Easy calculations show that, when $\gamma''' = 0$, the gross and net quantities are related by

$$\pi' = \frac{1}{1-\beta''}\left(\pi + \frac{\alpha' + \alpha''b}{\bar{a}_{x\,\overline{n}|}} + \beta' + \gamma' + \gamma''b\right), \tag{5.15}$$

and

$$V_t' = V_t - \frac{\bar{a}_{x+t\,\overline{n-t|}}}{\bar{a}_{x\,\overline{n|}}}(\alpha' + \alpha'' b). \tag{5.16}$$

It is seen that $\pi' > \pi$, as was anticipated at the outset. Furthermore, $V_t' < V_t$ for $0 \leq t < n$, which may be less obvious. The relationship (5.16) can be explained as follows: All expenses that incur at a constant rate throughout the term of the contract are compensated by an equal component in the “effective” gross premium rate $(1-\beta'')\pi'$, see (5.15). Thus, the only expense factor that appears in the gross reserve is the non-amortized initial α-cost, which is the last term on the right of (5.16). It represents a debt on the part of the insured and is therefore to be subtracted from the net reserve.

In Paragraph 4.3.D we have advocated non-negativity of the reserve. Now, already from (5.11) it is clear that the gross premium sets out negative at the time of issue of the contract and it will remain negative for some time thereafter until a sufficient amount of premium has been collected. The only way to get around this problem would be to charge an initial lump sum premium no less than the initial expense, but this is usually not done in practice (presumably) because a substantial down payment might deter customers with liquidity problems from buying insurance.

5.2 The general multi-state policy

A. General treatment of expenses. In Chapter 7 we are going to consider a more general multi-state insurance policy. Expenses are easily accommodated in the theory of that chapter since they can be treated as additional benefits of annuity and assurance type. Thus, from a technical point of view expenses do not create any additional difficulties, and we can therefore suitably end this chapter here. We round off by saying that expenses are still of conceptual and great practical importance. Assumptions about the various forms of expenses are part of the technical basis, which must be verified by the insurer and is subject to approval of the supervisory authority. Thus, just as statistical and economic analysis is required as a basis for assumptions about mortality and interest, thorough cost analysis are required as a basis for assumptions about the expense factors.

Chapter 6

Multi-life insurances

6.1 Insurances depending on the number of survivors

A. The single-life status reinterpreted. In the treatment of the single life status (x) in Chapters 3–4 we were having in mind the remaining life time T of an x year old person. From a mathematical point of view this interpretation is not essential. All that matters is that T is a non-negative random variable with an absolutely continuous distribution function, so that the survival function is of the form

$$ {}_tp_x = e^{-\int_0^t \mu_{x+\tau}\, d\tau} . \tag{6.1}$$

The footscript x serves merely to indicate what mortality law is in play. Regardless of the nature of the status (x) and the notion of lifetime represented by T, the previous results remain valid. In particular, all formulas for expected present values of payments depending on T are preserved, the basic ones being the endowment,

$$ {}_nE_x \quad = \quad v^n\, {}_np_x \,, \tag{6.2}$$

the life annuity,

$$ \bar{a}_{x\,\overline{n}|} = \int_0^n v^t\, {}_tp_x dt = \int_0^n {}_tE_x dt \,, \tag{6.3}$$

the endowment insurance,

$$ \bar{A}_{x\,\overline{n}|} = 1 - r\bar{a}_{x\,\overline{n}|} \,, \tag{6.4}$$

and the term insurance,

$$ \bar{A}^1_{x\,\overline{n}|} = \bar{A}_{x\,\overline{n}|} - {}_nE_x \,. \tag{6.5}$$

These formulas demonstrate that present values of all main types of payments in life insurance — endowments, life annuities, and assurances — can be traced back to the present value ${}_tE_x$ of an endowment and, as far as the mortality law is concerned, to the survival function ${}_tp_x$. Once we have determined ${}_tp_x$, all other functions of interest are obtained by integration, possibly by some numerical method, and elementary algebraic operations.

B. Multi-dimensional survival functions. Consider a body of r individuals, the j-th of which is called (x_j) and has remaining lifetime T_j, $j = 1, \ldots, r$. For the time being we shall confine ourselves to the case with independent lives. Thus, assume that the T_j are stochastically independent, and that each T_j possesses an intensity denoted by μ_{x_j+t} and, hence, has survival function

$$ {}_tp_{x_j} = e^{-\int_0^t \mu_{x_j+\tau}\, d\tau} . \tag{6.6}$$

(The function μ need not be the same for all j as the notation suggests; we have dropped an extra index j just to save notation.) The simultaneous distribution of $T_1, \ldots, T_r$ is given by the multi-dimensional survival function

$$\mathbb{P}\left[\cap_{j=1}^r \{T_j > t_j\}\right] = \prod_{j=1}^r {}_{t_j}p_{x_j} = e^{-\sum_{j=1}^r \int_0^{t_j} \mu_{x_j+\tau}\, d\tau}$$

or, equivalently, by the density

$$\prod_{j=1}^r {}_{t_j}p_{x_j} \mu_{x_j+t_j} . \tag{6.7}$$

C. The joint-life status. The *joint life* status $(x_1 \ldots x_r)$ is defined by having remaining lifetime

$$T_{x_1 \ldots x_r} = \min\{T_1, \ldots, T_r\} . \tag{6.8}$$

Thus, the r lives are looked upon as a single entity, which continues to exist as long as all members survive, and terminates upon the first death. The survival function of the joint-life is denoted by ${}_tp_{x_1 \ldots x_r}$ and is

$${}_tp_{x_1 \ldots x_r} = \mathbb{P}\left[\cap_{j=1}^r \{T_j > t\}\right] = e^{-\int_0^t \sum_{j=1}^r \mu_{x_j+\tau}\, d\tau} . \tag{6.9}$$

From this survival function we form the present values of an endowment ${}_nE_{x_1 \ldots x_r}$, a life annuity $\bar{a}_{x_1 \ldots x_r \,\overline{n|}}$, an endowment insurance $\bar{A}_{x_1 \ldots x_r \,\overline{n|}}$, and a term insurance, $\bar{A}^1_{x_1 \ldots x_r \,\overline{n|}}$, by just putting (6.9) in the role of the survival function in (6.2) – (6.5).

By inspection of (6.9), the mortality intensity of the joint-life status is simply the sum of the component mortality intensities,

$$\mu_{x_1 \ldots x_r}(t) = \sum_{j=1}^r \mu_{x_j+t} . \tag{6.10}$$

In particular, if the component lives are subject to G-M mortality laws with a common value of the parameter c,

$$\mu_{x_j+t} = \alpha_j + \beta_j e^{\gamma\,(x_j+t)}, \tag{6.11}$$

then (6.10) becomes

$$\mu_{x_1\ldots x_r}(t) = \alpha' + \beta' e^{\gamma\,t} \tag{6.12}$$

with

$$\alpha' = \sum_{j=1}^{r} \alpha_j\,,\ \ \beta' = \sum_{j=1}^{r} \beta_j e^{\gamma\,x_j}\,, \tag{6.13}$$

again a G-M law with the same γ as in the component laws.

D. The last-survivor status. The *last survivor* status $\overline{x_1\ldots x_r}$ is defined by having remaining lifetime

$$T_{\overline{x_1\ldots x_r}} = \max\{T_1,\ldots,T_r\}\,. \tag{6.14}$$

Now the r lives are looked upon as an entity that continues to exist as long as at least one member survives, and terminates upon the last death. The survival function of this status is denoted by ${}_tp_{\overline{x_1\ldots x_r}}$. By the general addition rule for probabilities (Appendix C),

$$\begin{aligned} {}_tp_{\overline{x_1\ldots x_r}} &= \mathbb{P}\left[\cup_{j=1}^{r}\{T_j > t\}\right] \\ &= \sum_{j} {}_tp_{x_j} - \sum_{j_1<j_2} {}_tp_{x_{j_1}x_{j_2}} + \ldots + (-1)^{r-1}\,{}_tp_{x_1\ldots x_r}\,. \end{aligned} \tag{6.15}$$

This way actuarial computations for the last survivor are reduced to computations for joint lives, which are simple. As explained in Paragraph A, all main types of present values can be built from (6.15). Formulas for benefits contingent on survival, obtained from (6.2) and (6.3), will reflect the structure of (6.15) in an obvious way. Formulas for death benefits are obtained from (6.4) and (6.5). The expressions are displayed in the more general case to be treated in the next paragraph.

E. The q survivors status.
The q *survivors* status $\overline{^{q}}_{x_1\ldots x_r}$ is defined by having as remaining lifetime the $(r-q+1)$-th order statistic of the sample $\{T_1,\ldots,T_r\}$. Thus the status is "alive" as long as there are at least q survivors among the original r. The survival function can be expressed in terms of joint life survival functions of sub-groups of lives by direct application of the theorem in Appendix C:

$${}_tp_{\overline{^{q}}_{x_1\ldots x_r}} = \sum_{p=q}^{r}(-1)^{p-q}\binom{p-1}{q-1}\sum_{j_1<\ldots<j_p} {}_tp_{x_{j_1}\ldots x_{j_p}}\,. \tag{6.16}$$

Present values of standard forms of insurances for the q survivors status are now obtained along the lines described in the previous paragraph. First, combine (6.16) with (6.2) to obtain

$$ {}_nE_{\overline{x_1\ldots x_r}^{\,q}} = \sum_{p=q}^{r} (-1)^{p-q} \binom{p-1}{q-1} \sum_{j_1<\ldots<j_p} {}_nE_{x_{j_1}\ldots x_{j_p}} \,, \tag{6.17} $$

the notation being self-explaining. Next, combine (6.3) and (6.17) to obtain

$$ \bar{a}_{\overline{x_1\ldots x_r}^{\,q}\,\overline{n}|} = \sum_{p=q}^{r} (-1)^{p-q} \binom{p-1}{q-1} \sum_{j_1<\ldots<j_p} \bar{a}_{x_{j_1}\ldots x_{j_p}\,\overline{n}|} \,. \tag{6.18} $$

Finally, present values of endowment and term insurances are obtained by inserting (6.18) and (6.17) in the general relations (6.4) and (6.5):

$$ \bar{A}_{\overline{x_1\ldots x_r}^{\,q}\,\overline{n}|} = 1 - r\bar{a}_{\overline{x_1\ldots x_r}^{\,q}\,\overline{n}|} \,, \tag{6.19} $$

$$ \bar{A}^1_{\overline{x_1\ldots x_r}^{\,q}\,\overline{n}|} = \bar{A}_{\overline{x_1\ldots x_r}^{\,q}\,\overline{n}|} - {}_nE_{x_{j_1}\ldots x_{j_p}} \,. \tag{6.20} $$

The following alternative to the expression in (6.19) has some aesthetic appeal as it expresses the insurance by corresponding insurances on joint lives:

$$ \bar{A}_{\overline{x_1\ldots x_r}^{\,q}\,\overline{n}|} = \sum_{p=q}^{r} (-1)^{p-q} \binom{p-1}{q-1} \sum_{j_1<\ldots<j_p} \bar{A}_{x_{j_1}\ldots x_{j_p}\,\overline{n}|} \,. \tag{6.21} $$

It is obtained upon substituting (6.18) on the right of (6.19), then inserting (recall (6.4)) $\bar{a}_{x_{j_1}\ldots x_{j_p}} = (1 - \bar{A}_{x_{j_1}\ldots x_{j_p}})/r$ and using (C.8) in Appendix C. A similar expression for the term insurance is obtained upon subtracting (6.17) from (6.21).

Chapter 7

Markov chains in life insurance

7.1 The insurance policy as a stochastic process

A. The basic entities. Consider an insurance policy issued at time 0 for a finite term of n years. We have in mind life or pension insurance or some other form of insurance of persons like disability or sickness coverage. In such lines of business benefits and premiums are typically contingent upon transitions of the policy between certain states specified in the contract. Thus, we assume there is a finite set of states, $\mathcal{Z} = \{0, 1, \ldots, J\}$, such that the policy at any time is in one and only one state, commencing in state 0 (say) at time 0. Denote the state of the policy at time t by $Z(t)$. Regarded as a function from $[0, n]$ to $\mathcal{Z}$, Z is assumed to be right-continuous, with a finite number of jumps, and $Z(0) = 0$. To account for the random course of the policy, Z is modeled as a stochastic process on some probability space $(\Omega, \mathcal{H}, \mathbb{P})$.

B. Model deliberations; realism versus simplicity. On specifying the probability model, two concerns must be kept in mind, and they are inevitably conflicting. On the one hand, the model should reflect the essential features of (a certain piece of) reality, and this speaks for a complex model to the extent that reality itself is complex. On the other hand, the model should be mathematically tractable, and this speaks for a simple model allowing of easy computation of quantities of interest. The art of modeling is to strike the right balance between these two concerns.

Favouring simplicity in the first place, we shall be working under Markov assumptions, which allow for fairly easy computation of relevant probabilities and expected values. Later on we shall demonstrate the versatility of this model framework, showing that it is capable of representing virtually any conception one might have of the mechanisms governing the development of the policy. We shall take the Markov chain model presented in [26] as a suitable framework throughout this text. A useful basic source is [30].

7.2 The continuous time Markov chain

A. The Markov property. A stochastic process is essentially determined by its finite-dimensional distributions. In the present case, where Z has only a finite state space, these are fully specified by the probabilities of the elementary events $\cap_{h=1}^{p}[Z(t_h) = j_h]$, $t_1 < \cdots < t_p$ in $[0, n]$ and $j_1, \ldots, j_p \in \mathcal{Z}$. Now, by the definition of conditional probability,

$$\begin{aligned}\mathbb{P}\left[Z(t_h) = j_h,\, h = 1, \ldots, p\right] \;&=\; \mathbb{P}\left[Z(t_p) = j_p \mid Z(t_g) = j_g,\, g = 0, \ldots, p-1\right] \times \\ &\;\; \mathbb{P}\left[Z(t_g) = j_g,\, g = 0, \ldots, p-1\right],\end{aligned}$$

and proceeding by induction we find

$$\begin{aligned}&\mathbb{P}\left[Z(t_h) = j_h,\, h = 1, \ldots, p\right] \\ &\quad = \prod_{h=1}^{p} \mathbb{P}\left[Z(t_h) = j_h \mid Z(t_g) = j_g,\, g = 0, \ldots, h-1\right]. \qquad (7.1)\end{aligned}$$

Here we have put $t_0 = 0$ and $j_0 = 0$ so that $[Z(t_0) = j_0]$ is the trivial event with probability 1. Thus, the specification of $\mathbb{P}$ could suitably start with the conditional probabilities appearing on the right of (7.1).

A particularly simple structure is obtained by assuming that, for all $t_1 < \cdots < t_h$ in $[0, n]$ and $j_1, \ldots, j_h \in \mathcal{Z}$,

$$\begin{aligned}&\mathbb{P}\left[Z(t_h) = j_h \mid Z(t_g) = j_g,\, g = 1, \ldots, h-1\right] \\ &\quad = \mathbb{P}[Z(t_h) = j_h \mid Z(t_{h-1}) = j_{h-1}], \qquad (7.2)\end{aligned}$$

which means that process is fully determined by the *(simple) transition probabilities*

$$p_{jk}(t, u) = \mathbb{P}[Z(u) = k \mid Z(t) = j], \qquad (7.3)$$

$t < u$ in $[0, n]$ and $j, k \in \mathcal{Z}$. In fact, if (7.2) holds, then (7.1) reduces to

$$\mathbb{P}\left[Z(t_h) = j_h,\, h = 1, \ldots, p\right] = \prod_{h=1}^{p} p_{j_{h-1} j_h}(t_{h-1}, t_h). \qquad (7.4)$$

One easily proves the equivalent that, for any $t_1 < \cdots < t_p < t < t_{p+1} < \cdots < t_{p+q}$ in $[0, n]$ and $j_1, \ldots, j_p, j, j_{p+1}, \ldots, j_{p+q}$ in $\mathcal{Z}$,

$$\begin{aligned}&\mathbb{P}\left[\, Z(t_h) = j_h,\, h = p+1, \ldots, p+q \mid Z(t) = j,\, Z(t_h) = j_h, h = 1, \ldots, p\right] \\ &\quad = \mathbb{P}\left[\, Z(t_h) = j_h,\, h = p+1, \ldots, p+q \mid Z(t) = j\right]. \qquad (7.5)\end{aligned}$$

Proclaiming t "the present time", (7.5) says that the future of the process is independent of its past when the present is known. (*Fully* known, that is; if the present state is only partly known, it may certainly help to add information about the past.)

The condition (7.2) is called the *Markov property.* We shall assume that Z possesses this property and, accordingly, call it a continuous time *Markov process* on the state space $\mathcal{Z}$.

From the simple transition probabilities we form the more general transition probability from j to some subset $\mathcal{K} \subset \mathcal{Z}$,

$$p_{j\mathcal{K}}(t,u) = \mathbb{P}[Z(u) \in \mathcal{K} \,|\, Z(t) = j] = \sum_{k \in \mathcal{K}} p_{jk}(t,u)\,. \tag{7.6}$$

We have, of course,

$$p_{j\mathcal{Z}}(t,u) = \sum_{k \in \mathcal{Z}} p_{jk}(t,u) = 1\,. \tag{7.7}$$

B. Alternative definitions of the Markov property. It is straightforward to demonstrate that (7.2), (7.4), and (7.5) are equivalent, so that any one of the three could have been taken as definition of the Markov property. Then (7.4) should be preceded by: "Assume there exist non-negative functions $p_{jk}(t,u)$, $j,k \in \mathcal{Z}$, $0 \le t \le u \le n$, such that $\sum_{k \in \mathcal{Z}} p_{jk}(t,u) = 1$ and, for any $0 \le t_1 < \cdots < t_p$ in $[0,n]$ and $\{j_1,\ldots,j_p\}$ in $\mathcal{Z}$,..."

We shall briefly outline more general definitions of the Markov property. For $\mathcal{T} \subset [0,n]$ let $\mathcal{H}_{\mathcal{T}}$ denote the class of all events generated by $\{Z(t)\}_{t \in \mathcal{T}}$. It represents everything that can be observed about Z in the time set $\mathcal{T}$. For instance, $\mathcal{H}_{\{t\}}$ is the information carried by the process at time t and consists of the elementary events $\emptyset$, Ω, and $[Z(t) = j]$, $j = 0,\ldots,J$, and all possible unions of these events. More generally, $\mathcal{H}_{\{t_1,\ldots,t_p\}}$ is the information carried by the process at times $t_1,\ldots,t_p$. Some sets $\mathcal{T}$ of interval type are frequently encountered, and we abbreviate $\mathcal{H}_{\le t} = \mathcal{H}_{[0,t]}$ (the entire history of the process by time t), $\mathcal{H}_{<t} = \mathcal{H}_{[0,t)}$ (the strict past of the process by time t), and $\mathcal{H}_{>t} = \mathcal{H}_{(t,n]}$ (the future of the process by time t).

The process Z is said to be a Markov process if, for any $B \in \mathcal{H}_{>t}$,

$$\mathbb{P}[B \,|\, \mathcal{H}_{\le t}] = \mathbb{P}[B \,|\, \mathcal{H}_{\{t\}}]\,. \tag{7.8}$$

This is the general form of (7.5).

An alternative definition says that, for any $A \in \mathcal{H}_{<t}$ and $B \in \mathcal{H}_{>t}$,

$$\mathbb{P}[A \cap B \,|\, \mathcal{H}_{\{t\}}] = \mathbb{P}[A \,|\, \mathcal{H}_{\{t\}}]\, \mathbb{P}[B \,|\, \mathcal{H}_{\{t\}}]\,, \tag{7.9}$$

that is, the past and the future of the process are conditionally independent, given its present state. In the case with finite state space (countability is equally simple) it is easy to prove that (7.8) and (7.9) are equivalent by working with the finite-dimensional distributions, that is, take $A \in \mathcal{H}_{\{t_1,\ldots,t_p\}}$ and $B \in \mathcal{H}_{\{t_{p+1},\ldots,t_{p+q}\}}$ with $t_1 < \cdots < t_p < t < t_{p+1} < \cdots < t_{p+q}$.

C. The Chapman-Kolmogorov equation. For a fixed $t \in [0,n]$ the events $\{Z(t) = j\}$, $j \in \mathcal{Z}$, are disjoint and their union is the almost sure event. It follows that

$$\mathbb{P}[Z(u) = k \,|\, Z(s) = i] = \sum_{j \in \mathcal{Z}} \mathbb{P}[Z(t) = j,\, Z(u) = k \,|\, Z(s) = i]$$
$$= \sum_{j \in \mathcal{Z}} \mathbb{P}[Z(t) = j \,|\, Z(s) = i]\, \mathbb{P}[Z(u) = k \,|\, Z(s) = i\,,\, Z(t) = j]\,.$$

If Z is Markov, and $0 \le s \le t \le u$, this reduces to

$$p_{ik}(s,u) = \sum_{j \in \mathcal{Z}} p_{ij}(s,t) p_{jk}(t,u)\,, \tag{7.10}$$

which is known as the *Chapman-Kolmogorov equation*.

D. Intensities of transition. In principle, specifying the Markov model amounts to specifying the $p_{jk}(t,u)$ in such a manner that the expressions on the right of (7.4) define probabilities in a consistent way. This would be easy if Z were a discrete time Markov chain with t ranging in a finite time set $0 = t_0 < t_1 < \cdots < t_q = n$: then we could just take the $p_{jk}(t_{q-1}, t_q)$ as any non-negative numbers satisfying $\sum_{k=0}^{r} p_{jk}(t_{p-1}, t_p) = 1$ for each $j \in \mathcal{Z}$ and $p = 1, \ldots, q$. This simple device does not carry over without modification to the continuous time case since there are no smallest finite time intervals from which we can build all probabilities by (7.4). An obvious way of adapting the basic idea to the time-continuous case is to add smoothness assumptions that give meaning to a notion of transition probabilities in infinitesimal time intervals.

More specifically, we shall assume that the *intensities of transition*,

$$\mu_{jk}(t) = \lim_{h \downarrow 0} \frac{p_{jk}(t, t+h)}{h} \tag{7.11}$$

exist for each $j, k \in \mathcal{Z}$, $j \neq k$, and $t \in [0,n)$ and, moreover, that they are piece-wise continuous. Another way of phrasing (7.11) is

$$p_{jk}(t, t+dt) = \mu_{jk}(t)dt + o(dt)\,, \tag{7.12}$$

where the term $o(dt)$ is such that $o(dt)/dt \to 0$ as $dt \to 0$. Thus, transition probabilities over a short time interval are assumed to be (approximately) proportional to the length of the interval, and the proportionality factors are just the intensities, which may depend on the time. What is "short" in this connection depends on the sizes of the intensities. For instance, if the $\mu_{jk}(\tau)$ are approximately constant and $<< 1$ for all $k \neq j$ and all $\tau \in [t, t+1]$, then $\mu_{jk}(t)$ approximates the transition probability $p_{jk}(t, t+1)$. In general, however, the intensities may attain any positive values and should not be confused with probabilities.

For $j \notin \mathcal{K} \subset \mathcal{Z}$, we define the intensity of transition from state j to the set of states $\mathcal{K}$ at time t as

$$\mu_{j\mathcal{K}}(t) = \lim_{u \downarrow t} \frac{p_{j\mathcal{K}}(t,u)}{u-t} = \sum_{k \in \mathcal{K}} \mu_{jk}(t)\,. \tag{7.13}$$

In particular, the total intensity of transition out of state j at time t is $\mu_{j,\mathcal{Z}-\{j\}}(t)$, which is abbreviated

$$\mu_{j\cdot}(t) = \sum_{k;k\neq j} \mu_{jk}(t)\,. \tag{7.14}$$

From (7.7) and (7.12) we get

$$p_{jj}(t, t+dt) = 1 - \mu_{j\cdot}(t)dt + o(dt)\,. \tag{7.15}$$

E. The Kolmogorov differential equations.
The transition probabilities are two-dimensional functions of time, and in non-trivial situations it is virtually impossible to specify them directly in a consistent manner or even figure how they should look on intuitive grounds. The intensities, however, are one-dimensional functions of time and, being easily interpretable, they form a natural starting point for specification of the model. Luckily, as we shall now see, they are also basic entities in the system as they determine the transition probabilities uniquely.

Suppose the process Z is in state j at time t. To find the probability that the process will be in state k at a given future time u, let us condition on what happens in the first small time interval $(t, t+dt]$. In the first place Z may remain in state j with probability $1 - \mu_{j\cdot}(t)\,dt$ and, conditional on this event, the probability of ending up in state k at time u is $p_{jk}(t+dt, u)$. In the second place, Z may jump to some other state g with probability $\mu_{jg}(t)\,dt$ and, conditional on this event, the probability of ending up in state k at time u is $p_{gk}(t+dt, u)$. Thus, the total probability of Z being in state k at time u is

$$\begin{aligned} p_{jk}(t,u) &= (1 - \mu_{j\cdot}(t)\,dt)\,p_{jk}(t+dt,u) \\ &\quad + \sum_{g;g\neq j} \mu_{jg}(t)\,dt\,p_{gk}(t+dt,u) + o(dt)\,, \end{aligned} \tag{7.16}$$

Upon putting $d_t p_{jk}(t,u) = p_{jk}(t+dt,u) - p_{jk}(t,u)$ in the infinitesimal sense, we arrive at

$$d_t p_{jk}(t,u) = \mu_{j\cdot}(t)\,dt\,p_{jk}(t,u) - \sum_{g;g\neq j} \mu_{jg}(t)\,dt\,p_{gk}(t,u)\,. \tag{7.17}$$

For given k and u these differential equations determine the functions $p_{jk}(\cdot, u)$, $j = 0, \ldots, J$, uniquely when combined with the obvious conditions

$$p_{jk}(u,u) = \delta_{jk}\,. \tag{7.18}$$

Here δ_{jk} is the Kronecker delta defined as 1 if $j = k$ and 0 otherwise.

The relation (7.16) could have been put up directly by use of the Chapman-Kolmogorov equation (7.10), with s, t, i, j replaced by $t, t + dt, j, g$, but we have carried through the detailed (still informal though) argument above since it will be in use repeatedly throughout the text. It is called the backward (differential) argument since it focuses on t, which in the perspective of the considered time period $[t, u]$ is the very beginning. Accordingly, (7.17) is referred to as the *Kolmogorov backward differential equations*, being due to A.N. Kolmogorov.

At points of continuity of the intensities we can divide by dt in (7.17) and obtain a limit on the right as dt tends to 0. Thus, at such points we can write (7.17) as

$$\frac{\partial}{\partial t} p_{jk}(t, u) = \mu_{j\cdot}(t) p_{jk}(t, u) - \sum_{g; g \neq j} \mu_{jg}(t) p_{gk}(t, u) \,. \tag{7.19}$$

Since we have assumed that the intensities are piece-wise continuous, the indicated derivatives exist piece-wise. We prefer, however, to work with the differential form (7.17) since it is generally valid under our assumptions and, moreover, invites algorithmic reasoning; numerical procedures for solving differential equations are based on approximation by difference equations for some fine discretization and, in fact, (7.16) is basically what one would use with some small $dt > 0$.

As one may have guessed, there exist also *Kolmogorov forward differential equations*. These are obtained by focusing on what happens at the end of the time interval in consideration. Reasoning along the lines above, we have

$$p_{ij}(s, t + dt) = \sum_{g; g \neq j} p_{ig}(s, t)\, \mu_{gj}(t)\, dt + p_{ij}(s, t)(1 - \mu_{j\cdot}(t)\, dt) + o(dt) \,,$$

hence

$$d_t p_{ij}(s, t) = \sum_{g; g \neq j} p_{ig}(s, t) \mu_{gj}(t)\, dt \; - p_{ij}(s, t) \mu_{j\cdot}(t)\, dt \,. \tag{7.20}$$

For given i and s, the differential equations (7.20) determine the functions $p_{ij}(s, \cdot)$, $j = 0, \ldots, J$, uniquely in conjunction with the obvious conditions

$$p_{ij}(s, s) = \delta_{ij} \,. \tag{7.21}$$

In some simple cases the differential equations have nice analytical solutions, but in most non-trivial cases they must be solved numerically, e.g. by the Runge-Kutta method.

Once the simple transition probabilities are determined, we may calculate the probability of any event in $\mathcal{H}_{\{t_1, \ldots, t_r\}}$ from the finite-dimensional distribution (7.4). In fact, with finite $\mathcal{Z}$ every such probability is just a finite sum of probabilities of elementary events to which we can apply (7.4).

Probabilities of more complex events that involve an infinite number of coordinates of Z, e.g. events in $\mathcal{H}_{\mathcal{T}}$ with $\mathcal{T}$ an interval, cannot in general be calculated from the simple transition probabilities. Often we can, however, put up differential equations for the requested probabilities and solve these by some suitable method.

Of particular interest is the probability of staying uninterruptedly in the current state for a certain period of time,

$$p_{\overline{jj}}(t,u) = \mathbb{P}[Z(\tau) = j, \tau \in (t,u] \,|\, Z(t) = j] \,. \tag{7.22}$$

Obviously $p_{\overline{jj}}(t,u) = p_{\overline{jj}}(t,s)\, p_{\overline{jj}}(s,u)$ for $t < s < u$. By the "backward" construction and (7.15) we get

$$p_{\overline{jj}}(t,u) = (1 - \mu_{j\cdot}(t)\, dt)\, p_{\overline{jj}}(t+dt,u) + o(dt) \,. \tag{7.23}$$

From here proceed as above, using $p_{\overline{jj}}(u,u) = 1$, to obtain

$$p_{\overline{jj}}(t,u) = e^{-\int_t^u \mu_{j\cdot}} \,. \tag{7.24}$$

F. Backward and forward integral equations. From the backward differential equations we obtain an equivalent set of integral equations as follows. Switch the first term on the right over to the left and, to obtain a complete differential there, multiply on both sides by the integrating factor $e^{\int_t^u \mu_{j\cdot}}$:

$$d_t\left(e^{\int_t^u \mu_{j\cdot}} p_{jk}(t,u)\right) = -e^{\int_t^u \mu_{j\cdot}} \sum_{g;g\neq j} \mu_{jg}(t)\, dt\, p_{gk}(t,u).$$

Now integrate over $(t,u]$ and use (7.18) to obtain

$$\delta_{jk} - e^{\int_t^u \mu_{j\cdot}} p_{jk}(t,u) = -\int_t^u e^{\int_\tau^u \mu_{j\cdot}} \sum_{g;\, g\neq j} \mu_{jg}(\tau) p_{gk}(\tau,u)\, d\tau \,.$$

Finally, carry the Kronecker delta over to the right, multiply by $-e^{-\int_t^u \mu_j}$, and use (7.24) to arrive at the backward integral equations

$$p_{jk}(t,u) = \int_t^u p_{\overline{jj}}(t,\tau) \sum_{g;\, g\neq j} \mu_{jg}(\tau) p_{gk}(\tau,u)\, d\tau + \delta_{jk} p_{\overline{jj}}(t,u) \,. \tag{7.25}$$

In a similar manner we obtain the following set of forward integral equations from (7.20):

$$p_{ij}(s,t) = \delta_{ij} p_{\overline{ii}}(s,t) + \sum_{g;\, g\neq j} \int_s^t p_{ig}(s,\tau) \mu_{gj}(\tau) p_{\overline{jj}}(\tau,t) d\tau \,. \tag{7.26}$$

The integral equations could be put up directly upon summing the probabilities of disjoint elementary events that constitute the event in question. For (7.26) the argument goes as follows. The first term on the right accounts for the possibility of ending up in state j without making any intermediate transitions, which is relevant only if $i = j$. The second term accounts for the possibility of ending up in state j after having made intermediate transitions and is the sum, over all states $g \neq j$ and all small time intervals $(\tau, \tau + d\tau)$ in (s,t), of the probability of arriving for the last time in state j from state g in the time interval $(\tau, \tau + d\tau)$. In a similar manner (7.25) is obtained upon splitting up by the direction and the time of the first departure, if any, from state j.

We now turn to some specializations of the model pertaining to insurance of persons. In each case we will choose such names and symbols for states and intensities as the situation suggests.

7.3 Examples

A. A single life with one cause of death. The life length of a person is modeled as a positive random variable T with survival function $\bar{F}$. There are two states, 'alive' and 'dead'. Labeling these by 0 and 1, respectively, the state process Z is simply

$$Z(t) = 1[T \leq t]\,,\ t \in [0,n]\,,$$

which counts the number of deaths by time $t \geq 0$. The process Z is right-continuous and is obviously Markov since in state 0 the past is trivial, and in state 1 the future is trivial. The transition probabilities are

$$p_{00}(s,t) = \bar{F}(t)/\bar{F}(s)\,.$$

The Chapman-Kolmogorov equation reduces to the trivial

$$p_{00}(s,u) = p_{00}(s,t)p_{00}(t,u)$$

or $\bar{F}(u)/\bar{F}(s) = \{\bar{F}(t)/\bar{F}(s)\}\{\bar{F}(u)/\bar{F}(t)\}$. The only non-null intensity is $\mu_{01}(t) = \mu(t)$, and

$$p_{00}(t,u) = e^{-\int_t^u \mu}\,. \tag{7.27}$$

The Kolmogorov differential equations reduce to just the definition of the intensity (write out the details).

The simple two state process with state 1 absorbing is outlined in Fig. 7.1

B. A single life with J causes of death. In the previous paragraph it was, admittedly, the process set-up that needed the example and not the other way around. The process formulation flexes its muscles only when we turn to more complex situations. Fig. 7.2 outlines a first extension of the model in the Paragraph A, whereby the single absorbing state ("dead") is replaced by J absorbing states representing different causes of death, e.g. "dead in accident", "dead from heart disease", etc. The index 0 in the intensities μ_{0j} is superfluous and has been dropped.

Relation (7.14) implies that the total mortality intensity is the sum of the intensities of death from different causes,

$$\mu(t) = \sum_{j=1}^{J} \mu_j(t)\,. \tag{7.28}$$

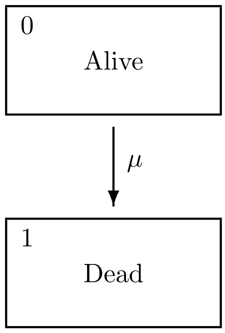


Figure 7.1: Sketch of the mortality model with one cause of death.

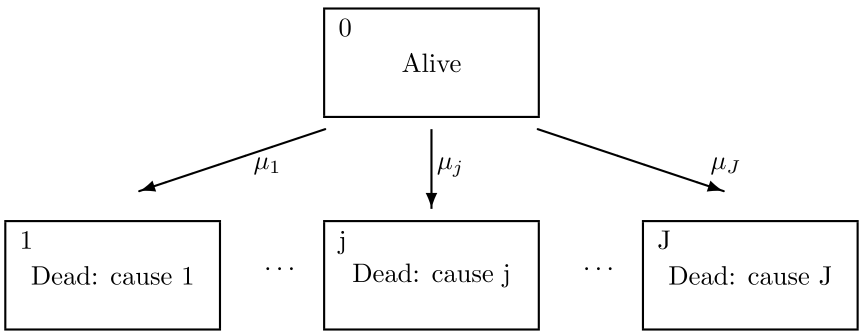


Figure 7.2: Sketch of the mortality model with J causes of death.

For a person aged t the probability of survival to u is the well-known survival probability $p_{00}(t,u)$ given by (7.27). The present enriched model opens possibilities of expressing ideas about the relative importance of various causes of death and thus better motivate a specific mortality law in the aggregate. For instance, the G-M law in the simple mortality model may be motivated as resulting from two causes of death, one with intensity α independent of age (pure accidents) and the other with intensity $\beta e^{\gamma t}$ (ageing and wear-out).

The probability that a t years old will die from cause j before age u is

$$p_{0j}(t,u) = \int_t^u e^{-\int_t^\tau \mu} \mu_j(\tau)\, d\tau\,. \tag{7.29}$$

This follows from e.g. (7.25) upon noting that $p_{jj}(t,u) = 1$, but – being totally transparent – it can be put up directly.

Inspection of (7.28) – (7.29) gives rise to a comment. An increase of one mortality intensity μ_k results in a decrease of the survival probability (evidently) and also of the probabilities of death from every other cause $j \neq k$, hence (since the probabilities sum to 1) an increase of the probability of death from cause k (also evident). Thus, the increased proportions of deaths from heart diseases and cancer in our times could be sufficiently explained by the fact that medical progress has practically eliminated mortality by lunge inflammation, childbed fever, and a number of other diseases.

The above discussion supports the assertion that the intensities are basic entities. They are the pure expressions of the forces acting on the policy in each given state, and the transition probabilities are resultants of the interplay between these forces.

C. A model for disabilities, recoveries, and death. Fig. 7.3 outlines a model suitable for analyzing insurances with payments depending on the state of health of the insured, e.g. sickness insurance providing an annuity benefit during periods of disability or life insurance with premium waiver during disability. Many other problems fit into the same scheme by mere re-labeling of the states. For instance, in connection with a pension insurance with additional benefits to the spouse, states a and i would be "unmarried" and "married", and in connection with unemployment insurance they would be "employed" and "unemployed".

For a person who is active at time s the Kolmogorov forward differential (7.20) equations are

$$\frac{\partial}{\partial t} p_{aa}(s,t) = p_{ai}(s,t)\rho(t) - p_{aa}(s,t)(\mu(t) + \sigma(t))\,, \tag{7.30}$$

$$\frac{\partial}{\partial t} p_{ai}(s,t) = p_{aa}(s,t)\sigma(t) - p_{ai}(s,t)(\nu(t) + \rho(t))\,. \tag{7.31}$$

(The probability $p_{ad}(s,t)$ is determined by the other two.) The initial conditions (7.21) become

$$p_{aa}(s,s) = 1\,, \tag{7.32}$$

$$p_{ai}(s,s) = 0\,. \tag{7.33}$$

(For a person who is disabled at time s the forward differential equations are the same, only with the first subscript a replaced by i in all the probabilities, and the side conditions are $p_{aa}(s,s) = 0\,,\;\; p_{ii}(s,s) = 1\,.$).

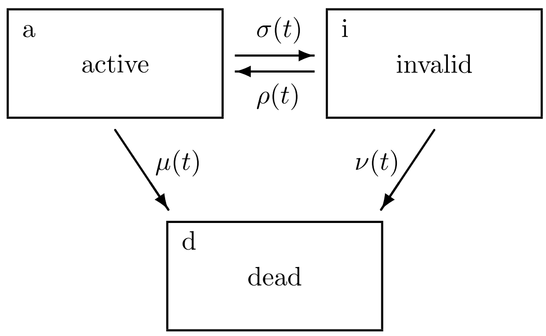


Figure 7.3: Sketch of a Markov chain model for disabilities, recoveries, and death.

When the intensities are sufficiently simple functions, one may find explicit closed expressions for the transition probabilities. Work through the case with constant intensities.

7.4 Selection phenomena

A. Introductory remarks. The Markov model (like any other model) may be accused of being oversimplified. For instance, in the disability model it says that the prospects of survival of a disabled person are unaffected by information about the past such as the pattern of previous disabilities and recoveries and, in particular, the duration since the last onset of disability. One could imagine that there are several types or degrees of disability, some of them light, with rather standard mortality, and some severe with heavy excess mortality. In these circumstances information about the past may be relevant: if we get to know that the onset of disability incurred a long time ago, then it is likely that one of the light forms is in play; if it incurred yesterday, it may well be one of the severe forms by which a soon death is to be expected.

The kind of heterogeneity effect mentioned here can be accommodated in the Markov framework by simply extending the state space, replacing the single state "disabled" by more states corresponding to different degrees of disability. From this Markov model we deduce the model of what we can observe as sketched in Fig. 7.3 upon letting the state "disabled" be the aggregate of the disability states in the basic model. What we end up with is typically no longer a Markov model.

Generally speaking, by variation of state space and intensities, the Markov set-up is capable of representing extremely complex phenomena. In the following we shall formalize these ideas with a particular view to selection phenomena often encountered in insurance. The ideas are to a great extent taken from [27].

Before proceeding we need to elaborate a bit on the concepts of transition intensity and Markov process and see how they are related. In the first place there are processes, Markov and others, that do not possess intensities (the reader is urged to invent some examples). In the second place, intensities may be possessed also by processes that are not Markov (although the definition (7.11) appears to imply that the concept of intensity is something that is based on the Markov assumption). For a general process Z with finite state space $\mathcal{Z}$ we can reasonably define the intensity of transition from state j to state k, given some additional past information $A \in \mathcal{H}_{<t}$, as

$$\mu_{jk}(t, A) = \lim_{dt \downarrow 0} \frac{\mathbb{P}[Z(t+dt) = k \,|\, Z(t) = j,\, A]}{dt}\,, \tag{7.34}$$

provided the limit exists. The process is Markov if and only if the intensity is independent of the past history,

$$\mu_{jk}(t, A) = \mu_{jk}(t)\,. \tag{7.35}$$

The 'only if' part follows from the previous definition of the Markov property, and the 'if' part follows from the Kolmogorov differential equations.

B. Aggregating states of a Markov chain. Let Z be the general continuous time Markov chain introduced in Section 7.2. Let $\{\mathcal{Z}_0, \ldots, \mathcal{Z}_{\tilde{J}}\}$ be a partition of $\mathcal{Z}$, that is, the $\mathcal{Z}_g$ are disjoint and their union is $\mathcal{Z}$. By convention, $0 \in \mathcal{Z}_0$. Define a stochastic process $\tilde{Z}$ on the state space $\tilde{\mathcal{Z}} = \{0, \ldots, \tilde{J}\}$ by

$$\tilde{Z}(t) = g \;\text{ iff }\; Z(t) \in \mathcal{Z}_g\,. \tag{7.36}$$

The interpretation is that we can observe the process $\tilde{Z}$ which represents summary information about some not fully observable Markov process Z.

Suppose the underlying process Z is observed to be in state i at time s. The subsequent development of $\tilde{Z}$ can be projected by conditional probabilities for the process Z. We have, for $s < t < u$,

$$\mathbb{P}[\tilde{Z}(u) = h \,|\, Z(s) = i,\, \tilde{Z}(t) = g] = \frac{1}{p_{i\mathcal{Z}_g}(s,t)} \sum_{j \in \mathcal{Z}_g} p_{ij}(s,t) p_{j\mathcal{Z}_h}(t,u)\,.$$

From this expression we obtain conditional transition intensities of the aggregate process:

$$\lim_{u \downarrow t} \frac{\mathbb{P}[\tilde{Z}(u) = h \,|\, Z(s) = i,\, \tilde{Z}(t) = g]}{u - t} = \frac{1}{p_{i\mathcal{Z}_g}(s,t)} \sum_{j \in \mathcal{Z}_g} p_{ij}(s,t) \mu_{j\mathcal{Z}_h}(t)\,.$$

We can not speak of *the* intensities since they would in general be different if more information about Z were conditioned on.

As an example, consider the aggregate of the states a and i in the disability model in Paragraph 7.3.C, $\mathcal{Z}_0 = \{a, i\}$, and put $\mathcal{Z}_1 = \{d\}$. Thus we observe only whether the insured is alive or not. The process $\tilde{Z}$ is Markov, of course (recall the argument in Paragraph 7.3.A). The survival probability is

$$\tilde{p}_{00}(0,t) = p_{aa}(0,t) + p_{ai}(0,t)\,,$$

and the mortality intensity at age t is

$$\tilde{\mu}(t) = \frac{p_{aa}(0,t)\mu(t) + p_{ai}(0,t)\nu(t)}{p_{aa}(0,t) + p_{ai}(0,t)}\,,$$

a weighted average of the mortality intensities of active and disabled, the weights being the probabilities of staying in the respective states.

C. Non-differential intensities. Suppose the transition intensities $\mu_{j\mathcal{Z}_h}(t)$ considered as functions of j are constant on each $\mathcal{Z}_g$, that is, there exist functions $\tilde{\mu}_{gh}(t)$ such that, for each t and $g, h \in \tilde{\mathcal{Z}}$,

$$\mu_{j\mathcal{Z}_h}(t) = \tilde{\mu}_{gh}(t)\,,\ \forall j \in \mathcal{Z}_g\,. \tag{7.37}$$

Then we shall say that the intensities of transition between the subsets $\mathcal{Z}_g$ are *non-differential* (within the individual subsets). The following result is evident on intuitive grounds, but never the less merits emphasis.

Theorem 1. *If the transition intensities of the process Z between the subsets $\mathcal{Z}_g$ are non-differential, then the process $\tilde{Z}$ defined by (7.36) is Markov with transition intensities $\tilde{\mu}_{gh}(t)$ defined by (7.37).*

Proof: Assume (7.37) is true. We must show that (7.35) holds for the process $\tilde{Z}$. Let A be an event depending only on $\{\tilde{Z}(\tau)\}_{0\le\tau<t}$. Using in successive order the facts that $[\tilde{Z}(t) = g] = \cup_{j\in\mathcal{Z}_g}[Z(t) = j]$ is a union of disjoint events, that $A \in \mathcal{H}_{<t}$, and that Z is Markov, we get

$$\begin{aligned}\mathbb{P}[A,\, \tilde{Z}(t) = g,\, \tilde{Z}(t+dt) = h] \;&=\; \sum_{j\in\mathcal{Z}_g} \mathbb{P}[A,\, Z(t) = j,\, Z(t+dt) \in \mathcal{Z}_h]\\ &= \sum_{j\in\mathcal{Z}_g} \mathbb{P}[A,\, Z(t) = j] p_{j\mathcal{Z}_h}(t, t+dt)\,.\\ &= \mathbb{P}[A,\, \tilde{Z}(t) = g]\, \tilde{\mu}_{gh}(t)\, dt + o(dt),\end{aligned}$$

which means that $\tilde{\mu}_{gh}(t, A) = \tilde{\mu}_{gh}(t)$ is independent of A. $\square$

D. Non-differential mortality. Let state J be absorbing, representing death, and let $\mathcal{H} = \{0, \ldots, J-1\}$ be the aggregate of states where the insured is alive. Assume that the mortality is non-differential, which means that all μ_{jJ}, $j \in \mathcal{H}$, are identical and equal to λ, say. Then, by Theorem 1, the survival probability is the same in all states $j \in \mathcal{H}$:

$$p_{j\mathcal{H}}(t,u) = e^{-\int_t^u \lambda}. \tag{7.38}$$

The conditional probability of staying in state $k \in \mathcal{H}$ at time t, given survival, is

$$p_{jk\,|\,\mathcal{H}}(t,u) = \frac{p_{jk}(t,u)}{p_{j\mathcal{H}}(t,u)} = p_{jk}(t,u)e^{\int_t^u \lambda}. \tag{7.39}$$

Inserting $p_{jk}(t,u) = p_{jk\,|\,\mathcal{H}}(t,u)e^{-\int_t^u \lambda}$ into (7.25), we get for each $j,k \in \mathcal{H}$ that

$$\begin{aligned} p_{jk\,|\,\mathcal{H}}(t,u)e^{-\int_t^u \lambda} \quad = \quad & \int_t^u e^{-\int_t^\tau \mu_{j,\mathcal{H}-\{j\}} - \int_t^\tau \lambda} \sum_{g \in \mathcal{H}-\{j\}} \mu_{jg}(\tau) p_{gk\,|\,\mathcal{H}}(\tau,u) e^{-\int_\tau^u \lambda} d\tau \\ & + \delta_{jk} e^{-\int_t^u \mu_{j,\mathcal{H}-\{j\}} - \int_t^u \lambda}. \end{aligned}$$

Multiplying with $e^{\int_s^t \lambda}$, we see that the conditional probabilities in (7.39) satisfy the integral equations (7.25) for the transition probabilities in the so-called *partial model* with state space $\mathcal{H}$ and transition intensities $\mu_{j,k}$, $j,k \in \mathcal{H}$. Thus, to find the transition probabilities in the full model, work first in the simple partial model for the states as alive and multiply the partial probabilities obtained there with the survival probability.

7.5 The standard multi-state contract

A. The contractual payments. We refer to the insurance policy with development as described in Paragraph 7.1.A. Taking Z to be a stochastic process with right-continuous paths and at most a finite number of jumps, the same holds also for the associated *indicator processes* I_j and *counting processes* N_{jk} defined, respectively, by $I_j(t) = 1[Z(t) = j]$ (1 or 0 according as the policy is in the state j or not at time t) and $N_{jk}(t) = \sharp\{\tau;\ Z(\tau-) = j,\ Z(\tau) = k,\ \tau \in (0,t]\}$ (the number of transitions from state j to state k ($k \neq j$) during the time interval $(0,t]$). The indicator processes $\{I_j(t)\}_{t \geq 0}$ and the counting processes $\{N_{jk}(t)\}_{t \geq 0}$ are related by the fact that I_j increases/decreases (by 1) upon a transition into/out of state j. Thus

$$dI_j(t) = dN_{\cdot j}(t) - dN_{j\cdot}(t), \tag{7.40}$$

where a dot in the place of a subscript signifies summation over that subscript, e.g. $N_{j\cdot} = \sum_{k;k \neq j} N_{jk}$.

The policy is assumed to be of standard type, which means that the payment function representing contractual benefits less premiums is of the form (recall the device (A.15))

$$dB(t) = \sum_k I_k(t)\,dB_k(t) + \sum_{k\neq\ell} b_{k\ell}(t)\,dN_{k\ell}(t)\,, \tag{7.41}$$

where each B_k, of form $dB_k(t) = b_k(t)\,dt + B_k(t) - B_k(t-)$, is a deterministic payment function specifying payments due during sojourns in state k (a general life annuity), and each $b_{k\ell}$ is a deterministic function specifying payments due upon transitions from state k to state ℓ (a general life assurance). When different from 0, $B_k(t) - B_k(t-)$ is an endowment at time t. The functions b_k and $b_{k\ell}$ are assumed to be finite-valued and piecewise continuous. The set of discontinuity points of any of the annuity functions B_k is $\mathcal{D} = \{t_0, t_1, \ldots, t_q\}$ (say).

Positive amounts represent benefits and negative amounts represent premiums. In practice premiums are only of annuity type. At times $t \notin [0, n]$ all payments are null.

B. Identities revisited. Here we make an intermission to make a comment that does not depend on the probability structure to be specified below. The identity (4.18) rests on the corresponding identity (4.17) between the present values. The latter is, in its turn, a special case of the identities put up in Section 2.1, from which many identities between present values in life insurance can be derived.

Suppose the investment portfolio of the insurance company bears interest with intensity $r(t)$ at time t. The following identity, which expresses life annuities by endowments and life assurances, is easily obtained upon integrating by parts, using (7.40):

$$\begin{aligned}\int_t^u e^{-\int_0^\tau r} I_j(\tau)\,dB_j(\tau) \quad &= \quad e^{-\int_0^u r} I_j(u)\,B_j(u) \; - \; e^{-\int_0^t r} I_j(t)B_j(t)\\ &\quad + \int_t^u e^{-\int_0^\tau r} I_j(\tau)B_j(\tau)r(\tau)\,d\tau\\ &\quad + \int_t^u e^{-\int_0^\tau r} B_j(\tau_-)\,d(N_{j\cdot}(\tau) - N_{\cdot j}(\tau))\,.\end{aligned}$$

C. Expected present values and prospective reserves. At any time $t \in [0, n]$, the present value of future benefits less premiums under the contract is

$$V(t) = \int_t^n e^{-\int_t^\tau r} dB(\tau)\,. \tag{7.42}$$

This is a liability for which the insurer is to provide a reserve, which by statute is the expected value. Suppose the policy is in state j at time t. Then the conditional expected value of $V(t)$ is

$$V_j(t) = \int_t^n e^{-\int_t^\tau r} \sum_k p_{jk}(t,\tau) \left(dB_k(\tau) + \sum_{\ell;\ell\neq k} b_{k\ell}(\tau)\mu_{k\ell}(\tau)\,d\tau \right)\,. \tag{7.43}$$

This follows by taking expectation under the integral in (7.42), inserting $dB(\tau)$ from (7.41), and using

$$\mathbb{E}[I_k(\tau) \,|\, Z(t) = j] = p_{jk}(t, \tau)\,,$$
$$\mathbb{E}[dN_{k\ell}(\tau) \,|\, Z(t) = j] = p_{jk}(t, \tau)\mu_{k\ell}(\tau)\, d\tau\,.$$

We expound the result as follows. With probability $p_{jk}(t, \tau)$ the policy stays in state k at time τ, and if this happens the life annuity provides the amount $dB_k(\tau)$ during a period of length $d\tau$ around τ. Thus, the expected present value at time t of this contingent payment is $p_{jk}(t, \tau)e^{-\int_t^\tau r}\, dB_k(\tau)$. With probability $p_{jk}(t, \tau)\, \mu_{k\ell}(\tau)\, d\tau$ the policy jumps from state k to state ℓ during a period of length $d\tau$ around τ, and if this happens the assurance provides the amount $b_{k\ell}(\tau)$. Thus, the expected present value at time t of this contingent payment is $p_{jk}(t, \tau)\, \mu_{k\ell}(\tau)\, d\tau\, e^{-\int_t^\tau r}\, b_{k\ell}(\tau)$. Summing over all future times and types of payments, we find the total given by (7.43).

Let $0 \le t < u < n$. Upon separating payments in $(t, u]$ and in $(u, n]$ on the right of (7.43), and using Chapman-Kolmogorov on the latter part, we obtain

$$\begin{aligned} V_j(t) = &\int_t^u e^{-\int_t^\tau r} \sum_k p_{jk}(t, \tau) \left(dB_k(\tau) + \sum_{\ell;\ell\neq k} b_{k\ell}(\tau)\mu_{k\ell}(\tau)\, d\tau \right) \\ &+ e^{-\int_t^u r} \sum_k p_{jk}(t, u)\, V_k(u). \end{aligned} \tag{7.44}$$

This expression is also immediately obtained upon conditioning on the state of the policy at time u.

Throughout the term of the policy the insurance company must currently maintain a reserve to meet future net liabilities in respect of the contract. By statute, if the policy is in state j at time t, then the company is to provide a reserve that is precisely $V_j(t)$. Accordingly, the functions V_j are called the *(state-wise) prospective reserves* of the policy. One may say that the principle of equivalence has been carried over to time t, now requiring expected balance between the amount currently reserved and the discounted future liabilities, given the information currently available. (Only the present state of the policy is relevant due to the Markov property and the simple memoryless payments under the standard contract).

D. The backward (Thiele's) differential equations. By letting u approach t in (7.44), we obtain a differential form that displays the dynamics of the reserves. In fact, we are going to derive a set of backward differential equations and, therefore, take the opportunity to apply the direct backward differential argument demonstrated and announced previously in Paragraph 7.2.E.

Thus, suppose the policy is in state j at time $t \notin \mathcal{D}$. Conditioning on what happens in a small time interval $(t, t + dt]$ (not intersecting $\mathcal{D}$) we write

$$V_j(t) = b_j(t)\, dt + \sum_{k;k\neq j} \mu_{jk}(t)\, dt\, b_{jk}(t)$$

$$+(1-\mu_{j\cdot}(t)\,dt)e^{-r(t)\,dt}\,V_j(t+dt)+\sum_{k;k\neq j}\mu_{jk}(t)\,dt\,e^{-r(t)\,dt}\,V_k(t+dt)\,.$$

Proceeding from here along the lines of the simple case in Section 4.4, we easily arrive at the *backward* or *Thiele's differential equations* for the state-wise prospective reserves,

$$\begin{aligned}\frac{d}{dt}V_j(t) &= (r(t)+\mu_{j\cdot}(t))\,V_j(t)-\sum_{k;k\neq j}\mu_{jk}(t)\,V_k(t)\\ &\quad -b_j(t)-\sum_{k;k\neq j}b_{jk}(t)\mu_{jk}(t)\,.\end{aligned} \tag{7.45}$$

The differential equations are valid in the open intervals (t_{p-1},t_p), $p=1,\ldots,q$, and together with the conditions

$$V_j(t_p-)=(B_j(t_p)-B_j(t_p-))+V_j(t_p)\,,\quad p=1,\ldots,q,\,j\in\mathcal{Z}, \tag{7.46}$$

they determine the functions V_j uniquely.

A comment is in order on the differentiability of the V_j. At points of continuity of the functions b_j, b_{jk}, μ_{jk}, and r there is no problem since there the integrand on the right of (7.43) is continuous. At possible points of discontinuity of the integrand the derivative $\frac{d}{dt}V_j$ does not exist. However, since such discontinuities are finite in number, they will not affect the integrations involved in numerical procedures. Thus we shall throughout allow ourselves to write the differential equations on the form (7.45) instead of the generally valid differential form obtained upon putting $dV_j(t)$ on the left and multiplying with dt on the right.

E. Solving the differential equations. Only in rare cases of no practical interest is it possible to find closed form solutions to the differential equations. In practice one must resort to numerical methods to determine the prospective reserves. As a matter of experience a fourth order Runge-Kutta procedure works reliably in virtually all situations encountered in practice.

One solves the differential equations 'from top down'. First solve (7.45) in the upper interval (t_{q-1},n) subject to (7.46), which specializes to $V_j(n-)=B_j(n)-B_j(n-)$ since $V_j(n)=0$ for all j by definition. Then go to the interval below and solve (7.45) subject to $V_j(t_{q-1}-)=(B_j(t_{q-1})-B_j(t_{q-1}-))+V_j(t_{q-1})$, where $V_j(t_{q-1})$ was determined in the first step. Proceed in this manner downwards.

It is realized that the Kolmogorov backward equations (7.17) are a special case of the Thiele equations (7.45); the transition probability $p_{jk}(t,u)$ is just the prospective reserve in state j at time t for the simple contract with the only payment being a lump sum payment of 1 at time u if the policy is then in state k, and with no interest. Thus a numerical procedure for computation of prospective reserves can also be used for computation of the transition probabilities.

F. The equivalence principle. If the equivalence principle is invoked, one must require that

$$V_0(0) = -B_0(0)\,. \tag{7.47}$$

This condition imposes a constraint on the contractual functions b_j, B_j, and b_{jk}, viz. on the premium level for given benefits and 'design' of the premium plan. It is of a different nature than the conditions (7.46), which follow by the very definition of prospective reserves (for given contractual functions).

G. Savings premium and risk premium. The equation (7.45) can be recast as

$$-b_j(t)\,dt = dV_j(t) - r(t)\,dt\,V_j(t) + \sum_{k;k\neq j} R_{jk}(t)\mu_{jk}(t)\,dt\,. \tag{7.48}$$

where

$$R_{jk}(t) = b_{jk}(t) + V_k(t) - V_j(t)\,. \tag{7.49}$$

The quantity $R_{jk}(t)$ is called the *sum at risk* associated with (a possible) transition from state j to state k at time t since, upon such a transition, the insurer must immediately pay out the sum insured and also provide the appropriate reserve in the new state, but he can cash the reserve in the old state. Thus, the last term in (7.48) is the expected net outlay in connection with a possible transition out of the current state j in $(t, t+dt)$, and it is called the *risk premium*. The two first terms on the right of (7.48) constitute the *savings premium* in $(t, t+dt)$, called so because it is the amount that has to be provided to maintain the reserve in the current state; the increment of the reserve less the interest earned on it. On the left of (7.48) is the premium paid in $(t, t+dt)$, and so the relation shows how the premium decomposes in a savings part and a risk part. Although helpful as an interpretation, this consideration alone cannot carry the full understanding of the differential equation since (7.48) is valid also if $b_j(t)$ is positive (a benefit) or 0.

H. Integral equations. In (7.45) let us switch the term $(r(t) + \mu_{j\cdot}(t))\,V_j(t)$ appearing on the right of over to the left, and multiply the equation by $e^{-\int_0^t (r+\mu_{j\cdot})}$ to form a complete differential on the left:

$$\begin{aligned}
&\frac{d}{dt}\left(e^{-\int_0^t (r+\mu_{j\cdot})} V_j(t)\right) = \\
&\quad -e^{-\int_0^t (r+\mu_{j\cdot})}\left(\sum_{k;k\neq j} \mu_{jk}(t)\,V_k(t) + b_j(t) + \sum_{k;k\neq j} b_{jk}(t)\mu_{jk}(t)\right)\,.
\end{aligned}$$

Now integrate over an interval (t,u) containing no jumps $B_j(\tau)-B_j(\tau-)$ and, recalling that $e^{-\int_t^\tau \mu_{j\cdot}} = p_{\overline{jj}}(t,\tau)$, rearrange a bit to obtain the integral equation

$$\begin{aligned} V_j(t) \;&=\; \int_t^u p_{\overline{jj}}(t,\tau)\, e^{-\int_t^\tau r} \left(b_j(\tau) + \sum_{k;\, k\neq j} \mu_{jk}(\tau)(b_{jk}(\tau)+V_k(\tau)) \right) d\tau \\ &\quad + p_{\overline{jj}}(t,u)\, e^{-\int_t^u r}\, V_j(u-)\,. \end{aligned} \tag{7.50}$$

This result generalizes the backward integral equations for the transition probabilities (7.25) and, just as in that special case, also the expression on the right hand side of (7.50) is easy to interpret; it decomposes the future payments into those that fall due before and those that fall due after the time of the first transition out of the current state in the time interval (t,u) or, if no transition takes place, those that fall due before and those that fall due after time u.

We shall take a direct route to the integral equation (7.50) that actually is the rigorous version of the backward technique. Suppose that the policy is in state j at time t. Let us apply the rule of iterated expectations to the expected value $V_j(t)$, conditioning on whether a transition out of state j takes place within time u or not and, in case it does, also condition on the time and the direction of the first transition. We then get

$$\begin{aligned} V_j(t) \;&=\; \int_t^u p_{\overline{jj}}(t,\tau) \sum_{k;\, k\neq j} \mu_{jk}(\tau)\, d\tau \left(\int_t^\tau e^{-\int_t^s r} b_j(s)\, ds + e^{-\int_t^\tau r}(b_{jk}(\tau)+V_k(\tau)) \right) \\ &\quad + p_{\overline{jj}}(t,u) \left(\int_t^u e^{-\int_t^s r} b_j(s)\, ds + e^{-\int_t^u r}\, V_j(u-) \right). \end{aligned} \tag{7.51}$$

To see that this is the same as (7.50), we need only to observe that

$$\begin{aligned} &\int_t^u p_{\overline{jj}}(t,\tau)\, \mu_{j\cdot}(\tau) \int_t^\tau e^{-\int_t^s r} b_j(s)\, ds\, d\tau \\ &= \int_t^u \int_s^u \frac{d}{d\tau}(-p_{\overline{jj}}(t,\tau))\, d\tau e^{-\int_t^s r} b_j(s)\, ds \\ &= -p_{\overline{jj}}(t,u) \int_t^u e^{-\int_t^s r} b_j(s)\, ds + \int_t^u p_{\overline{jj}}(t,s) e^{-\int_t^s r} b_j(s)\, ds. \end{aligned}$$

I. Uses of the differential equations. If the contractual functions do not depend on the reserves, the defining relation (7.43) give explicit expressions for the state-wise reserves and strictly speaking the differential equations (7.45) are not needed for constructive purposes. They are, however, computationally convenient since there are good methods for numerical solution of differential equations. They also serve to give insight into the dynamics of the policy.

The situation is entirely different if the contractual functions are allowed to depend on the reserves in some way or other. The most typical examples are repayment of a part of the reserve upon withdrawal (a state "withdrawn" must then be included in the state space $\mathcal{Z}$) and expenses depending partly on the

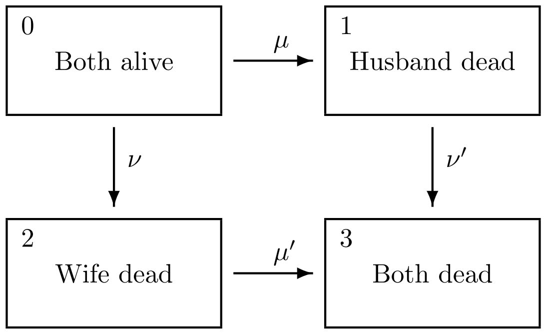


Figure 7.4: Sketch of a model for two lives.

reserve. Also the primary insurance benefits may in some cases be specified as functions of the reserve. In such situations the differential equations are an indispensable tool in the construction of the reserves and determination of the equivalence premium. We shall provide an example in the next paragraph.

J. An example: Widow's pension. A married couple buys a combined life insurance and widow's pension policy specifying that premiums are payable at level rate c as long as both husband and wife are alive, widow's pension is payable at level rate b (as long as the wife survives the husband), and a life assurance with sum s is due immediately upon the death of the husband if the wife is already dead (a benefit to their dependents). The policy terminates at time n. The relevant Markov model is sketched Fig. 7.4. Assume that the interest rate r is constant.

The differential equations (7.45) now specialize to the following (we omit the trivial equation for $V_3(t) = 0$):

$$\frac{d}{dt}V_0(t) = (r + \mu(t) + \nu(t))\, V_0(t) + c - \mu(t)V_1(t) - \nu(t)V_2(t)\,, \quad (7.52)$$

$$\frac{d}{dt}V_1(t) = (r + \nu'(t))\, V_1(t) - b\,, \quad (7.53)$$

$$\frac{d}{dt}V_2(t) = (r + \mu'(t))\, V_2(t) - \mu'(t)\, s\,. \quad (7.54)$$

Consider a modified contract under which 50% of the reserve is paid to the husband in the event of the wife's earlier death before time n, the philosophy being that couples receiving no pensions should have some of their savings back. Now the differential equations are really needed. Under the modified contract the equations above remain unchanged except that the term $0.5V_0(t)\nu(t)$ must be subtracted on the right of (7.52), which then changes to

$$\frac{d}{dt}V_0(t) = (r + \mu(t) + 0.5\nu(t))\, V_0(t) + c - \mu(t)V_1(t) - \nu(t)V_2(t)\,, \qquad (7.55)$$

Together with the conditions $V_j(n) = 0,\ j = 0, 1, 2$, these equations are easily solved.

As a second case the widow's pension shall be analyzed in the presence of administration expenses that depend partly on the reserve. Consider again the policy terms described in the introduction of this paragraph, but assume that administration expenses incur with an intensity that is a times the current reserve throughout the entire period $[0, n]$.

The differential equations for the reserves remain as in (7.52)–(7.54), except that for each j the term $a\,V_j(t)$ is to be subtracted on the right of the differential equation for V_j. Thus, the administration costs related to the reserve has the same effect as a decrease of the interest intensity r by a.

7.6 Select mortality revisited

A. A simple Markov chain model. Referring to Section 3.4 we shall present a simple Markov model that offers an explanation of the selection phenomenon.

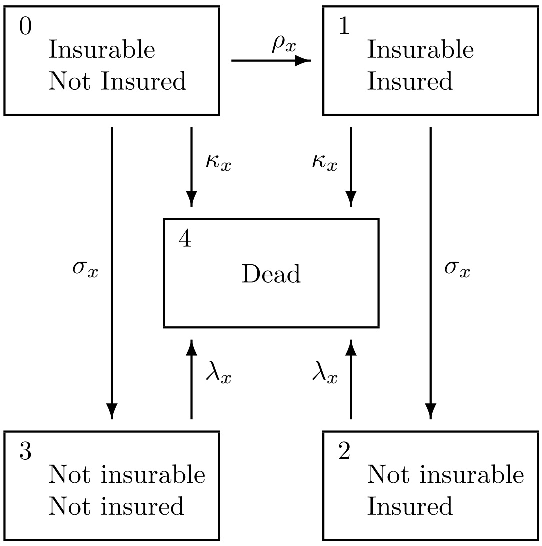


Figure 7.5: A Markov model for occurrences of non-insurability, purchase of insurance, and death.

The Markov model sketched in Fig. 7.5 is designed for studies of selection effects due to underwriting standards. The population is grouped into four categories or states by the criteria insurable/uninsurable and insured/not insured. In addition there is a category comprising the dead. It is assumed that each person enters state 0 as new-born and thereafter changes states in accordance with a time-continuous Markov chain with age-dependent forces of transition as indicated in the figure. Non-insurability occurs upon onset of disability or serious illness or other intervening circumstances that entail excess mortality. Hence it is assumed that

$$\lambda_x > \kappa_x ; \;\; x > 0. \tag{7.56}$$

Let $Z(x)$ be the state at age x for a randomly chosen new-born, and denote the transition probabilities of the Markov process $\{Z(x);\, x > 0\}$. The following formulas can be put up directly:

$$p_{11}(x, x+t) = \exp\{-\int_x^{x+t} (\sigma + \kappa)\}, \tag{7.57}$$

$$p_{12}(x, x+t) = \int_x^{x+t} \exp\{-\int_x^u (\sigma + \kappa)\}\, \sigma_u \exp\{-\int_u^{x+t} \lambda\}\, du, \tag{7.58}$$

$$p_{00}(0, x) = \exp\{-\int_0^x (\sigma + \kappa + \rho)\}\,. \tag{7.59}$$

B. Select mortality among insured lives. The insured lives are in either state 1 or state 2. Those who are in state 2 reached to buy insurance before they turned non-insurable. However, the insurance company does not observe transitions from state 1 to state 2; the only available information are x and $x+t$. Thus, the relevant survival function is

$${}_t p_{[x]} = p_{11}(x, x+t) + p_{12}(x, x+t)\,, \tag{7.60}$$

the probability that a person who entered state 1 at age x, will attain age $x+t$. The symbol on the left of (7.60) is chosen in accordance with standard actuarial notation, see Sections 3.3 – 3.4.

The force of mortality corresponding to (7.60) is

$$\mu_{[x]+t} = \frac{\kappa_{x+t}\, p_{11}(x, x+t) + \lambda_{x+t}\, p_{12}(x, x+t)}{p_{11}(x, x+t) + p_{12}(x, x+t)}. \tag{7.61}$$

(Self-evident by conditioning on $Z(x+t)$.) In general, the expression on the right of (7.61) depends effectively on both x and t, that is, mortality is select.

We can now actually establish that under the present model the select mortality intensity behaves as stated in Paragraph 3.4.C. It is suitable in the following to fix $x + t = y$, say, as we are interested in how the mortality at a certain age depends on the age of entry.

C. The select force of mortality is an decreasing function of the age at entry. Formula (7.61) can be recast as

$$\mu_{[x]+y-x} = \kappa_y + \zeta(x,y)(\lambda_y - \kappa_y)\,, \tag{7.62}$$

where

$$\zeta(x,y) = \frac{p_{12}(x,y)}{p_{11}(x,y)+p_{12}(x,y)} \tag{7.63}$$

$$= \frac{1}{1+p_{11}(x,y)/p_{12}(x,y)}\,. \tag{7.64}$$

We easily find that

$$p_{12}(x,y)/p_{11}(x,y) = \int_x^y \sigma_u \exp\{\int_u^y (\sigma+\kappa-\lambda)\}\,du\,,$$

which is a decreasing function of x. It follows that $\mu_{[x]+y-x}$ is a decreasing function of x as asserted in the heading of this paragraph.

The explanation is simple. Formula (7.61) expresses $\mu_{[x]+y-x}$ as a weighted average of κ_y and λ_y, the weights being (of course) the conditional probabilities of being insurable and non-insurable, respectively. The weight attached to λ_y, the larger of the two rates, decreases as x increases. Or, put in terms of everyday speech: in a body of insured lives of the same age x and duration $t = y - x$, some will have turned non-insurable in the period since entry; the longer the duration, the larger the proportion of non-insurable lives. In particular, those who have just entered, are known to be insurable, that is, $\mu_{[x]} = \kappa_x$.

D. Comparison with the mortality in the population. Let $\bar{\mu}_x$ denote the force of mortality of a randomly chosen life of age x from the population. A formula for $\bar{\mu}_x$ is easily obtained starting from the survival function ${}_x\bar{p}_0 = \sum_{i=0}^{3} p_{0i}(0,x)$. It can, however, also be picked directly from the results of the previous paragraph by noting that the pattern of mortality must be the same in the population as among lives insured as newly-born, i.e. $\bar{\mu}_y = \mu_{[0]+y}$. Then, since $\mu_{[x]+y-x}$ is a decreasing function of x and $\bar{\mu}_y$ corresponds to $x = 0$, it follows that $\bar{\mu}_y > \mu_{[x]+y-x}$ for all $x < y$.

Again the explanation is trivial; due to the underwriting standards, the proportion of non-insurable lives will be less among insured people than in the population as a whole.

7.7 Higher order moments of present values

A. Differential equations for moments of present values. Our framework is the Markov model and the standard insurance contract. The set of time points with possible lump sum annuity payments is $\mathcal{D} = \{t_0, t_1 \ldots, t_m\}$ (with $t_0 = 0$ and $t_m = n$).

Denote by $V(t,u)$ the present value at time t of the payments under the contract during the time interval $(t,u]$ and abbreviate $V(t) = V(t,n)$ (the present value at time t of all future payments). We want to determine higher order moments of $V(t)$. By the Markov property, we need only the state-wise conditional moments

$$V_j^{(q)}(t) = \mathbb{E}[V(t)^q | Z(t) = j]\,,$$

$q = 1, 2, \ldots$

Theorem 2. *The functions* $V_j^{(q)}$ *are determined by the differential equations*

$$\begin{aligned}\frac{d}{dt}V_j^{(q)}(t) &= (qr(t) + \mu_{j\cdot}(t))V_j^{(q)}(t) - qb_j(t)V_j^{(q-1)}(t) \\ &\quad - \sum_{k;\,k\neq j} \mu_{jk}(t) \sum_{p=0}^{q} \binom{q}{p} (b_{jk}(t))^p V_k^{(q-p)}(t)\,,\end{aligned}$$

valid on $(0,n)\backslash\mathcal{D}$ *and subject to the conditions*

$$V_j^{(q)}(t-) = \sum_{p=0}^{q} \binom{q}{p} (B_j(t) - B_j(t-))^p V_j^{(q-p)}(t)\,, \tag{7.65}$$

$t \in \mathcal{D}$. $\square$

Proof: Obviously, for $t < u < n$,

$$V(t) = V(t,u) + e^{-\int_t^u r} V(u)\,, \tag{7.66}$$

For any $q = 1, 2, \ldots$ we have by the binomial formula

$$V^q(t) = \sum_{p=0}^{q} \binom{q}{p} V(t,u)^p \left(e^{-\int_t^u r} V(u)\right)^{q-p}\,. \tag{7.67}$$

Consider first a small time interval $(t, t+dt]$ without any lump sum annuity payment. Putting $u = t + dt$ in (7.67) and taking conditional expectation, given $Z(t) = j$, we get

$$V_j^{(q)}(t) = \sum_{p=0}^{q} \binom{q}{p} \mathbb{E}\left[V(t,t+dt)^p \left(e^{-r(t)\,dt} V(t+dt)\right)^{q-p} \middle|\, Z(t) = j \right]. \tag{7.68}$$

By use of iterated expectations, conditioning on what happens in the small interval $(t, t+dt]$, the p-th term on the right of (7.68) becomes

$$\binom{q}{p} (1 - \mu_{j\cdot}(t)\,dt)\,(b_j(t)\,dt)^p\, e^{-(q-p)r(t)\,dt}\, V_j^{(q-p)}(t+dt) \tag{7.69}$$

$$+ \binom{q}{p} \sum_{k;\,k\neq j} \mu_{jk}(t)\,dt\,(b_j(t)\,dt + b_{jk}(t))^p e^{-(q-p)r(t)\,dt} V_k^{(q-p)}(t+dt)\,. \tag{7.70}$$

Let us identify the significant parts of this expression, disregarding terms of order $o(dt)$. First look at (7.69); for $p = 0$ it is

$$(1 - \mu_{j\cdot}(t)\,dt)e^{-qr(t)\,dt}\,V_j^{(q)}(t + dt)\,,$$

for $p = 1$ it is

$$q\,b_j(t)\,dt\,e^{-(q-1)r(t)\,dt}\,V_j^{(q-1)}(t + dt)\,,$$

and for $p > 1$ is $o(dt)$. Next look at (7.70); the factor

$$dt\,(b_j(t)\,dt + b_{jk}(t))^p = dt \sum_{r=0}^{p} \binom{p}{r} (b_j(t)\,dt)^r (b_{jk}(t))^{p-r}$$

reduces to $dt\,(b_{jk}(t))^p$ so that (7.70) reduces to

$$\binom{q}{p} \sum_{k;\,k\neq j} \mu_{jk}(t)\,dt\,(b_{jk}(t))^p e^{-(q-p)r(t)\,dt} V_k^{(q-p)}(t + dt)\,.$$

Thus, we gather

$$\begin{aligned} V_j^{(q)}(t) \quad &= \quad (1 - \mu_{j\cdot}(t)\,dt)e^{-qr(t)\,dt}\,V_j^{(q)}(t + dt) \\ &\quad + q\,b_j(t)\,dt\,e^{-(q-1)r(t)\,dt}\,V_j^{(q-1)}(t + dt) \\ &\quad + \sum_{p=0}^{q} \binom{q}{p} \sum_{k;\,k\neq j} \mu_{jk}(t)\,dt\,(b_{jk}(t))^p e^{-(q-p)r(t)\,dt} V_k^{(q-p)}(t + dt)\,. \end{aligned}$$

Now subtract $V_j^{(q)}(t + dt)$ on both sides, divide by dt, let dt tend to 0, and use $\lim_{t\downarrow 0}\left(e^{-qr(t)\,dt} - 1\right)/dt = -qr(t)$ to obtain the differential equation (7.65).

The condition (7.65) follows easily by putting $t - dt$ and t in the roles of t and u in (7.67) and letting dt tend to 0. □

A rigorous proof is given in [38].

Central moments are easier to interpret and therefore more useful than the non-central moments. Letting $m_t^{(q)j}$ denote the q-th central moment corresponding to the non-central $V_t^{(q)j}$, we have

$$m_j^{(1)}(t) \quad = \quad V_j^{(1)}(t)\,, \tag{7.71}$$

$$m_j^{(q)}(t) \quad = \quad \sum_{p=0}^{q} (-1)^{q-p} \binom{q}{p} V_j^{(p)}(t) \left(V_j^{(1)}(t)\right)^{q-p}\,. \tag{7.72}$$

B. Computations. The computation goes as follows. First solve the differential equations in the upper interval (t_{m-1}, n), where the side conditions (7.65) are just

$$V_j^{(q)}(n-) = (B_j(n) - B_j(n-))^q \tag{7.73}$$

since $V_j^{(q)}(n) = \delta_{q0}$ (the Kronecker delta). Then, if $m > 1$, solve the differential equations in the interval (t_{m-2}, t_{m-1}) subject to (7.65) with $t = t_{m-1}$, and proceed in this manner downwards.

C. Numerical examples. We shall calculate the first three moments for some standard forms of insurance related to the 'disability model' in Paragraph 7.3.C. We assume that the interest rate is constant and 4.5% per year,

$$r = \ln(1.045) = 0.044017\,,$$

and that the intensities of transitions between the states depend only on the age x of the insured and are

$$\begin{aligned}\mu_x &= \nu_x = 0.0005 + 0.000075858 \cdot 10^{0.038x}\,,\\ \sigma_x &= 0.0004 + 0.0000034674 \cdot 10^{0.06x}\,,\\ \rho_x &= 0.005\,.\end{aligned}$$

The intensities μ, ν, and σ are those specified in the G82M technical basis. (That basis does not allow for recoveries and uses $\rho = 0$).

Consider a male insured at age 30 for a period of 30 years, hence use $\mu_{02}(t) = \mu_{12}(t) = \mu_{30+t}$, $\mu_{01}(t) = \sigma_{30+t}$, $\mu_{10}(t) = \rho_{30+t}$, $0 < t < 30$ $(= n)$. The central moments $m_t^{(q)j}$ defined in (7.71) – (7.72) have been computed for the states 0 and 1 (state 2 is uninteresting) at times $t = 0, 6, 12, 18, 24$, and are shown
– in Table 7.1 for a term insurance with sum 1 $(= b_{02} = b_{12})$;
– in Table 7.2 for an annuity payable in active state with level intensity 1 $(= b_0)$;
– in Table 7.3 for an annuity payable in disabled state with level intensity 1 $(= b_1)$;
– in Table 7.4 for a combined policy providing a term insurance with sum 1 $(= b_{02} = b_{12})$ and a disability annuity with level intensity 0.5 $(= b_1)$ against level net premium 0.013108 $(= -b_0)$ payable in active state.
You should try to interpret the results.

D. Solvency margins in life insurance – an illustration. Let Y be the present value of all future net liabilities in respect of an insurance portfolio. Denote the q-th central moment of Y by $m^{(q)}$. The so-called normal power approximation of the upper ε-fractile of the distribution of Y, which we denote by $y_{1-\varepsilon}$, is based on the first three moments and is

$$y_{1-\varepsilon} \approx m^{(1)} + c_{1-\varepsilon}\sqrt{m^{(2)}} + \frac{c_{1-\varepsilon}^2 - 1}{6}\frac{m^{(3)}}{m^{(2)}}\,,$$

where $c_{1-\varepsilon}$ is the upper ε-fractile of the standard normal distribution. Adopting the so-called break-up criterion in solvency control, $y_{1-\varepsilon}$ can be taken as a

Table 7.1: Moments for a life assurance with sum 1

Time t	0	6	12	18	24	30
$m_t^{(1)0} = m_t^{(1)1}$:	0.0683	0.0771	0.0828	0.0801	0.0592	0
$m_t^{(2)0} = m_t^{(2)1}$:	0.0300	0.0389	0.0484	0.0549	0.0484	0
$m_t^{(3)0} = m_t^{(3)1}$:	0.0139	0.0191	0.0262	0.0343	0.0369	0

Table 7.2: Moments for an annuity of 1 per year while active:

Time t	0	6	12	18	24	30
$m_t^{(1)0}$:	15.763	13.921	11.606	8.698	4.995	0
$m_t^{(1)1}$:	0.863	0.648	0.431	0.230	0.070	0
$m_t^{(2)0}$:	5.885	5.665	4.740	2.950	0.833	0
$m_t^{(2)1}$:	7.795	5.372	3.104	1.290	0.234	0
$m_t^{(3)0}$:	−51.550	−44.570	−32.020	−15.650	−2.737	0
$m_t^{(3)1}$:	78.888	49.950	25.099	8.143	0.876	0

Table 7.3: Moments for an annuity of 1 per year while disabled:

Time t	0	6	12	18	24	30
$m_t^{(1)0}$:	0.277	0.293	0.289	0.239	0.119	0
$m_t^{(1)1}$:	15.176	13.566	11.464	8.708	5.044	0
$m_t^{(2)0}$:	1.750	1.791	1.646	1.147	0.364	0
$m_t^{(2)1}$:	11.502	8.987	6.111	3.107	0.716	0
$m_t^{(3)0}$:	15.960	14.835	11.929	6.601	1.277	0
$m_t^{(3)1}$:	−101.500	−71.990	−42.500	−17.160	−2.452	0

Table 7.4: Moments for a life assurance of 1 plus a disability annuity of 0.5 per year against net premium of 0.013108 per year while active:

Time t	0	6	12	18	24	30
$m_t^{(1)0}$:	0.0000	0.0410	0.0751	0.0858	0.0533	0
$m_t^{(1)1}$:	7.6451	6.8519	5.8091	4.4312	2.5803	0
$m_t^{(2)0}$:	0.4869	0.5046	0.4746	0.3514	0.1430	0
$m_t^{(2)1}$:	2.7010	2.0164	1.2764	0.5704	0.0974	0
$m_t^{(3)0}$:	2.1047	1.9440	1.5563	0.8686	0.1956	0
$m_t^{(3)1}$:	-12.1200	-8.1340	-4.3960	-1.5100	-0.1430	0

minimum requirement on the technical reserve at the time of consideration. It decomposes into the premium reserve, $m^{(1)}$, and what can be termed the fluctuation reserve, $y_{1-\varepsilon} - m^{(1)}$. A possible measure of the riskiness of the portfolio is the ratio $R = \left(y_{1-\varepsilon} - m^{(1)}\right)/P$, where P is some suitable measure of the size of the portfolio at the time of consideration. By way of illustration, consider a portfolio of N independent policies, all identical to the one described in connection with Table 7.4 and issued at the same time. Taking as P the total premium income per year, the value of R at the time of issue is 48.61 for $N = 10$, 12.00 for $N = 100$, 3.46 for $N = 1000$, 1.06 for $N = 10000$, and 0.332 for $N = 100000$.

7.8 A Markov chain interest model

A. The force of interest process.
The economy (or rather the part of the economy that governs the interest) is a homogeneous time-continuous Markov chain Y on a finite state space $\mathcal{Y} = \{1, \ldots, J^Y\}$, with intensities of transition λ_{ef}, $e, f \in \mathcal{Y}$, $e \neq f$. The force of interest is r_e when the economy is in state e, that is,

$$r(t) = \sum_e I_e^Y(t) r_e \,, \tag{7.74}$$

where $I_e^Y(t) = 1[Y(t) = e]$ is the indicator of the event that Y is in state e at time t.

Figure 7.6 shows a flow-chart of a simple Markov chain interest rate model with three states, 0.02, 0.05, 0.08. Direct transition can only be made to a neighbouring state, and the total intensity of transition out of any state is 0.5, that is, the interest rate changes every two years on the average. By symmetry, the long run average interest rate is 0.05.

B. The payment process.
We adopt the standard Markov chain model of a life insurance policy in Section

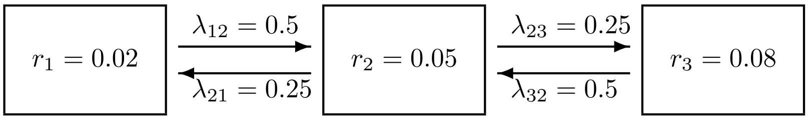


Figure 7.6: Sketch of a simple Markov chain interest model.

7.2 and equip the associated indicator and counting processes with topscript Z to distinguish them from the corresponding entities for the Markov chain governing the interest. We assume the payment stream is of the standard type considered in Section 7.5.

C. The full Markov model.
We assume that the processes Y and Z are independent. Then $X = (Y, Z)$ is a Markov chain on $\mathcal{X} = \mathcal{Y} \times \mathcal{Z}$ with intensities

$$\kappa_{ej,fk}(t) = \begin{cases} \lambda_{ef}(t)\,, & e \neq f,\, j = k\,, \\ \mu_{jk}(t)\,, & e = f,\, j \neq k\,, \\ 0\,, & e \neq f,\, j \neq k\,. \end{cases}$$

D. Reserves and higher order moments in the combined model.
For the purpose of assessing the contractual liability we are interested in aspects of its conditional distribution, given the available information at time t. We focus here on determining the conditional moments. By the Markov assumption, the functions in quest are the state-wise conditional moments

$$V_{ej}^{(q)}(t) = \mathrm{E}\left[\left(\int_t^n e^{-\int_t^\tau r}\, dB(\tau) \right)^q \,\middle|\, Y(t) = e, Z(t) = j \right].$$

Copying the proof in Section 7.7, we find that the functions $V_{ej}^{(q)}(\cdot)$ are determined by the differential equations

$$\begin{aligned} \frac{d}{dt} V_{ej}^{(q)}(t) &= (qr_e + \mu_{j\cdot}(t) + \lambda_{e\cdot}) V_{ej}^{(q)}(t) - q b_j(t) V_{ej}^{(q-1)}(t) \\ &- \sum_{k;k\neq j} \mu_{jk}(t) \sum_{p=0}^{q} \binom{q}{p} (b_{jk}(t))^p V_{ek}^{(q-p)}(t) - \sum_{f;f\neq e} \lambda_{ef} V_{fj}^{(q)}(t)\,, \end{aligned} \tag{7.75}$$

valid on $(0, n)\backslash\mathcal{D}$ and subject to the conditions

$$V_{ej}^{(q)}(t-) = \sum_{p=0}^{q} \binom{q}{p} (\Delta B_j(t))^p V_{ej}^{(q-p)}(t)\,,\ \ t \in \mathcal{D}\,.\ \Box \tag{7.76}$$

For $q = 2, 3, \ldots$, denote by $m_{ej}^{(q)}(t)$ the q-th central moment corresponding to $V_{ej}^{(q)}(t)$, and define $m_{ej}^{(1)}(t) = V_{ej}^{(1)}(t)$. Having computed the non-central moments, we obtain the central moments of orders $q > 1$ from

$$m_{ej}^{(q)}(t) = \sum_{p=0}^{q} \binom{q}{p} (-1)^{q-p} V_{ej}^{(p)}(t) \left(V_{ej}^{(1)}(t) \right)^{q-p} .$$

E. Numerical results for a combined insurance policy.
Consider a combined life insurance and disability pension policy issued at time 0 to a person who is then aged x, say. The relevant states of the policy are $a =$ *active*, $i =$ *disabled*, and $d =$ *dead*. At time t, when the insured is $x + t$ years old, transitions between these states take place with intensities

$$\begin{aligned}
\mu_{ad}(t) &= \mu_{id}(t) = 0.0005 + 0.000075858 \cdot 10^{0.038(x+t)} , \\
\mu_{ai}(t) &= 0.0004 + 0.0000034674 \cdot 10^{0.06(x+t)} , \\
\mu_{ia}(t) &= 0.005 .
\end{aligned}$$

We extend the model by assuming that the force of interest may assume three values, $r_1 = \ln(1.00) = 0$ (*low* – in fact no interest), $r_2 = \ln(1.045) = 0.04402$ (*medium*), and $r_3 = \ln(1.09) = 0.08618$ (*high*), and that the transitions between these states are governed by a Markov chain with infinitesimal matrix of the form

$$\Lambda = \lambda \begin{pmatrix} -1 & 1 & 0 \\ 0.5 & -1 & 0.5 \\ 0 & 1 & -1 \end{pmatrix} . \tag{7.77}$$

The scalar λ can be interpreted as the expected number of transitions per time unit and is thus a measure of interest rate volatility.

Table 1 displays the first three central moments of the present value at time 0 for the following case, henceforth referred to as *the combined policy* for short: the age at entry is $x = 30$, the term of the policy is $n = 30$, the benefits are a life assurance with sum 1 ($= b_{ad} = b_{id}$) and a disability annuity with level intensity 0.5 ($= b_i$), and premiums are payable in active state continuously at level rate π ($= -b_a$), which is taken to be the net premium rate in state (2,a) (i.e. the rate that establishes expected balance between discounted premiums and benefits when the insured is active and the interest is at medium level at time 0).

The first three rows in the body of the table form a benchmark; $\lambda = 0$ means no interest fluctuation, and we therefore obtain the results for three cases of fixed interest. It is seen that the second and third order moments of the present value are strongly dependent on the (fixed) force of interest and, in fact, their absolute values decrease when the force of interest increases (as could be expected since increasing interest means decreasing discount factors and, hence, decreasing present values of future amounts).

It is seen that, as λ increases, the differences across the three pairs of columns get smaller and in the end they vanish completely. The obvious interpretation

Table 7.5: Central moments $m_{ej}^{(q)}(0)$ of orders $q = 1, 2, 3$ of the present value of future benefits less premiums for *the combined policy* in interest state e and policy state j at time 0, for some different values of the rate of interest changes, λ. Second column gives the net premium π of a policy starting from interest state 2 (medium) and policy state 1 (active).

	e, j :		$1, a$	$1, i$	$2, a$	$2, i$	$3, a$	$3, i$
λ	π	q						
		1	0.15	13.39	0.00	7.65	−0.39	5.03
0	.0131	2	2.55	12.50	0.49	2.70	0.13	0.80
		3	20.45	−99.02	2.11	−12.12	0.37	−2.38
		1	0.06	11.31	0.00	7.90	−0.03	5.78
.05	.0137	2	1.61	12.26	0.62	5.41	0.25	2.43
		3	11.94	−42.87	3.20	−4.33	0.94	−0.08
		1	0.02	8.43	0.00	7.81	−0.02	7.24
.5	.0134	2	0.65	4.90	0.55	4.15	0.46	3.52
		3	3.34	−13.35	2.59	−10.13	2.02	−7.74
		1	0.00	7.77	0.00	7.70	0.00	7.64
5	.0132	2	0.51	2.86	0.50	2.91	0.49	2.86
		3	2.26	−12.51	2.20	−12.19	2.14	−11.88
		1	0.00	7.69	0.00	7.69	0.00	7.69
∞	.0132	2	0.50	2.74	0.50	2.74	0.50	2.74
		3	2.15	−12.37	2.15	−12.37	2.15	−12.37

is that the initial interest level is of little importance if the interest changes rapidly.

The overall impression from the two central columns corresponding to medium interest is that, as λ increases from 0, the variance of the present value will first increase to a maximum and then decrease again and stabilize. This observation supports the following piece of intuition: the introduction of moderate interest fluctuation adds uncertainty to the final result of the contract, but if the interest changes sufficiently rapidly, it will behave like fixed interest at the mean level. Presumably, the values of the net premium in the second column reflect the same effect.

F. Complement on Markov chains.
Let $X = \{X(t)\}_{t\geq 0}$ be a time-continuous Markov chain on the finite state space $\mathcal{X} = \{1, \ldots, J\}$. Denote by $P(t,u)$ the $J \times J$ matrix whose j,k-element is the transition probability $p_{jk}(t,u) = P[X(u) = k \mid X(t) = j]$. The Markov property implies the Chapman-Kolmogorov equation

$$P(s,u) = P(s,t)P(t,u)\,, \tag{7.78}$$

valid for $0 \leq s \leq t \leq u$. In particular

$$P(t,t) = I^{J\times J}\,, \tag{7.79}$$

the $J \times J$ identity matrix. The intensity of transition from state j to state k ($\neq j$) at time t is defined as $\kappa_{jk}(t) = \lim_{dt\downarrow 0} p_{jk}(t, t+dt)/dt$ or, equivalently, by

$$p_{jk}(t,t+dt) = \kappa_{jk}(t)\,dt + o(dt)\,, \tag{7.80}$$

when the limit exists. Then, obviously,

$$p_{jj}(t,t+dt) = 1 - \kappa_{j\cdot}(t)dt + o(dt)\,, \tag{7.81}$$

where $\kappa_{j\cdot}(t) = \sum_{k;\,k\neq j}\ \kappa_{jk}(t)$ can appropriately be termed the total intensity of transition out of state j at time t. The infinitesimal matrix $M(t)$ is the $J \times J$ matrix with $\kappa_{jk}(t)$ in row j and column k, defining $\kappa_{jj}(t) = -\kappa_{j\cdot}(t)$. With this notation (7.80) – (7.81) can be assembled in

$$P(t,t+dt) = I + M(t)dt\,. \tag{7.82}$$

The probabilities determine the intensities. Conversely, the probabilities are determined by the intensities through Kolmogorov's differential equations, which are readily obtained upon combining (7.78) and (7.82). There is a forward equation,

$$\frac{\partial}{\partial t}P(s,t) = P(s,t)M(t)\,, \tag{7.83}$$

and a backward equation,

$$\frac{\partial}{\partial t}P(t,u) = -M(t)P(t,u)\,, \tag{7.84}$$

each of which determine $P(t,u)$ when combined with the condition (7.79).

When $M(t) = M$, a constant, then (as is obvious from the Kolmogorov equations) $P(s,t) = P(0,t-s)$ depends on s and t only through $t-s$. In this case we write $P(t) = P(0,t)$, allowing a slight abuse of notation. The equations (7.83) – (7.84) now reduce to

$$\frac{d}{dt}P(t) = P(t)M = MP(t)\,. \tag{7.85}$$

The limit $\Pi = \lim_{t\to\infty} P(t)$ exists, and the j-th row of Π is the limiting (stationary) distribution of the state of the process, given that it starts from state j. We shall assume throughout that all states communicate with each other. Then the stationary distribution $\pi' = (\pi_1,\ldots,\pi_J)$, say, is independent of the initial state, and so

$$\Pi = 1^{J\times 1}\pi'\,, \tag{7.86}$$

where $1^{J\times 1}$ is the J-dimensional column vector with all entries equal to 1.

Letting $t\to\infty$ in (7.85) and using (7.86), we get $1^{J\times 1}\pi' M = M 1^{J\times 1}\pi' = 0^{J\times J}$ (a matrix of the indicated dimension with all elements equal to 0), that is,

$$\pi' M = 0^{1\times J}\,,\ M1^{J\times 1} = 0^{J\times 1}\,. \tag{7.87}$$

Thus, 0 is an eigenvalue of M, and π' and $1^{J\times 1}$ are corresponding left and right eigenvectors, respectively.

From Paragraph 4.3 of Karlin and Taylor (1975) we gather the following useful representation result. Let ρ_j, $j = 1,\ldots,J$, be the eigenvalues of M and, for each j, let ψ_j' and ϕ_j be the corresponding left and right eigenvectors, respectively. Let Φ be the $J\times J$ matrix whose j-th column is ϕ_j. Then the j-th row of Φ^{-1} is just ψ_j', and introducing $R(t) = \mathrm{diag}(e^{\rho_j t})$, the transition matrix $P(t)$ can be expressed as

$$P(t) = \Phi R(t)\Phi^{-1} = \sum_{j=1}^{J} e^{\rho_j t}\phi_j\psi_j'\,, \tag{7.88}$$

which is computationally convenient. We can take ρ_1 to be 0 and $\phi_1 = 1^{J\times 1}$. Then $\psi_1' = \pi'$, and we obtain

$$P(t) = 1^{J\times 1}\pi' + \sum_{j=2}^{J} e^{\rho_j t}\phi_j\psi_j'\,. \tag{7.89}$$

All the ρ_j, $j = 2,\ldots,J$ are strictly negative, and so the representation shows that the transition probabilities converge exponentially to the stationary distribution.

At this point we need to make precise that in (7.85) the $\frac{d}{dt}$ is to be thought of as an operator, to be distinguished from the matrix $P^{(1)}(t)$ of derivatives it produces when applied to $P(t)$. Now, for $\lambda > 0$, define

$$P_\lambda(t) = P(\lambda t)\,. \tag{7.90}$$

Upon differentiating this relationship and using (7.85), we obtain

$$\frac{d}{dt}P_\lambda(t) = \frac{d}{dt}P(\lambda t) = P^{(1)}(\lambda t)\lambda = P_\lambda(t)\lambda M\,,$$

which shows that $P_\lambda(t)$, which is certainly a matrix of transition probabilities, has infinitesimal matrix

$$M_\lambda = \lambda M\,. \tag{7.91}$$

Thus, doubling (say) the intensities of transition affects the transition probabilities the same way as a doubling of the time period.

7.9 Dependent lives

A. Introduction.
Actuarial tables for multi-life statuses are invariably based on the assumption of mutual independence between the remaining lengths of the individual component lives. The independence hypothesis is computationally convenient or, rather, was so in those days when tables had to be constructed. In the present era of scientific computing such concerns are not so important.

Let S and T be real-valued random variables defined on some probability space. Being mainly interested in survival analysis related to life insurance, we shall let S and T represent the remaining life lengths of two individuals insured under the same policy, let us say husband and wife, respectively. Thus, we make the convenient (but not essential) assumptions that S and T are strictly positive with probability 1 and that they possess a joint density.

The variables S and T are stochastically independent if

$$\mathbb{P}[S > s\,,\, T > t] \;=\; \mathbb{P}[S > s]\,\mathbb{P}[T > t] \text{ for all } s \text{ and } t.$$

In particular, stochastic independence implies that $\mathbb{C}(g(S), h(T)) = 0$ for all functions g and h such that the covariance is well defined. (We let $\mathbb{C}$ and $\mathbb{V}$ denote covariance and variance, respectively.)

Mortality statistics suggest that life lengths of husband and wife are dependent and, moreover, that they are positively correlated. It is easy to think of possible explanations to this empirical fact. For instance, that people who marry do so because they have something in common ('birds of a feather fly together'), or that married people share lifestyle and living conditions and therefore also hazards of diseases and accidents, or that death of the spouse impairs the living conditions for the survivor ('a grief effect', or maybe the husband just does not know where the kitchen is and starves to death shortly after the loss of the spouse).

Correlation is a rather special measure of dependence – essentially it measures linear dependence between random variables – and it is not sufficiently refined for our purposes.

B. Notions of positive dependence.
There are various notions of positive dependence between pairs of random variables, and we will introduce three of them here. A comprehensive reference text is [5].

Definition PQD: S and T are *positively quadrant dependent*, written $\mathrm{PQD}(S, T)$, if

$$\mathbb{P}[S > s\,,\, T > t] \,\geq\, \mathbb{P}[S > s]\,\mathbb{P}[T > t] \text{ for all } s \text{ and } t. \tag{7.92}$$

This definition is symmetric in the two variables, so $\mathrm{PQD}(S, T)$ is the same as $\mathrm{PQD}(T, S)$. The defining inequality (7.92) is equivalent to

$$\mathbb{P}[S > s \,|\, T > t] \,\geq\, \mathbb{P}[S > s]\,, \tag{7.93}$$

which is easy to interpret: knowing e.g. that the wife will survive a certain period years improves the survival prospects of the husband.

Definition AS: S and T are *associated*, written $\mathrm{AS}(S, T)$, if

$$\mathbb{C}(g(S, T), h(S, T)) \geq 0 \tag{7.94}$$

for all real-valued functions g and h that are increasing in both arguments (and for which the covariance exists).

Also the definition of AS is symmetric in the two variables, so $\mathrm{AS}(S, T)$ is the same as $\mathrm{AS}(T, S)$.

Definition RTI: S is *right tail increasing in* T, written $\mathrm{RTI}(S|T)$, if

$$\mathbb{P}[S > s \,|\, T > t] \text{ is an increasing function of } t \text{ for each fixed } s. \tag{7.95}$$

The definition of RTI is not symmetric in the two variables.

To each notion of positive dependence there is a corresponding notion of negative dependence. We can reasonably say that S and T are negatively quadrant dependent if the inequality (7.92) is reversed. This is the same as $\mathrm{PQD}(-S, T)$. We can say that S and T are negatively associated ('dissociated' does not have the right connotation) if the inequality (7.94) is reversed. This is the same as $\mathrm{AS}(-S, T)$. We say that S is *right tail decreasing in* T, written $\mathrm{RTD}(S|T)$ if $\mathbb{P}[S > s \,|\, T > t]$ is a decreasing function of t for each fixed s. This is the same as $\mathrm{RTD}(-S|T)$. Since results on positive dependence thus translate into results on negative dependence, we will henceforth focus on the former.

Theorem 1: $\mathrm{RTI}(S|T) \Rightarrow \mathrm{AS}(S, T) \Rightarrow \mathrm{PQD}(S, T)$.

Proof (incomplete): The first implication, $\mathrm{RTI}(S|T) \Rightarrow \mathrm{AS}(S, T)$, is the hard part. The proof is long and technical and can be found in [20].

The second implication $\mathrm{AS}(S,T) \Rightarrow \mathrm{PQD}(S,T)$ is easy. For $g(S,T) = 1_{(s,\infty)}(S) = 1[S > s]$ and $h(S,T) = 1_{(t,\infty)}(T) = 1[T > t]$, (7.94) reduces to

$$\mathbb{C}(1[S > s], 1[T > t]) \geq 0\,, \tag{7.96}$$

which is just a reformulation of the defining inequality (7.92).

As a partial compensation for the absence of proof of the first implication, let us prove the shortcut implication $\mathrm{RTI}(S|T) \Rightarrow \mathrm{PQD}(S,T)$: if $\mathrm{RTI}(S|T)$, then $\mathbb{P}[S > s \,|\, T > t] \geq \mathbb{P}[S > s \,|\, T > 0]$ for $t > 0$, which is the same as (7.93). $\square$

The following result is a partial converse to the second implication in Theorem 1. It could be formulated by saying that positive quadrant dependence is equivalent to "marginal association".

Lemma 1: $\mathrm{PQD}(S,T) \Leftrightarrow \mathbb{C}(g(S), h(T)) \geq 0$ for increasing functions g and h.

Proof: By (7.96) the result holds for (increasing) indicator functions $g(S) = 1[S > s]$ and $h(T) = 1[T > t]$. Then it holds for increasing simple functions, $g(S) = g_0 + \sum_{i=1}^{m} g_i 1[S > s_i]$ and $h(T) = h_0 + \sum_{j=1}^{n} h_j 1[T > t_j]$ with positive constant coefficients g_i and h_j as is seen from

$$\mathbb{C}(g(S), h(T)) = \sum_{i=1}^{m}\sum_{j=1}^{n} g_i h_j \mathbb{C}\left(1[S > s_i], 1[T > t_j]\right) \geq 0\,.$$

Then it holds for all increasing functions $g(S)$ and $h(T)$ since since any increasing function from $\mathbb{R}$ to $\mathbb{R}$ can be written as the limit of a sequence of increasing simple functions (monotone convergence). $\square$

In the definitions of PQD, AS, and RTI we could equally reasonably have entered the events $S \geq s$ and $T \geq t$. Due to the continuity property of probability measures, it does not matter which inequalities we use, $>$ or $\geq$. %See Exercise 21. For $A_1 \subseteq A_2 \subseteq \ldots$ we have $\lim \mathbb{P}[A_n] = \mathbb{P}[\cup_n A_n]$, and for $A_1 \supseteq A_2 \supseteq \ldots$ we have $\lim \mathbb{P}[A_n] = \mathbb{P}[\cap_n A_n]$.

C. Dependencies between present values.

The following table lists formulas for the life lengths and the survival functions of the four statuses husband, wife, their joint life, and their last survivor.

Status (z)	Life length U	Survival function $\mathbb{P}[U > \tau]$
Husband (x)	S	$\mathbb{P}[S > \tau]$
Wife (y)	T	$\mathbb{P}[T > \tau]$
Joint life (x, y)	$S \wedge T$	$\mathbb{P}[S > \tau\,,\, T > \tau]$
Last survivor $\overline{x, y}$	$S \vee T$	$\mathbb{P}[S > \tau] + \mathbb{P}[T > \tau] - \mathbb{P}[S > \tau\,,\, T > \tau]$

The next table recapitulate the formulas for present values and their expected values for the most basic insurance benefits to a status (z) with remaining life length U.

Payment scheme	Present value	Expected present value
Pure endowment	$e^{-rn}\,1[U > n]$	${}_nE_z = e^{-rn}\,\mathbb{P}[U > n]$
Life annuity	$\int_0^n e^{-r\tau}\,1[U > \tau]\,d\tau$	$\bar{a}_{z\,\overline{n}\rceil} = \int_0^n e^{-r\tau}\,\mathbb{P}[U > \tau]\,d\tau$
Term insurance	$e^{-rU}1[U \leq n]$	$\bar{A}_{\overset{1}{x}\,\overline{n}\rceil} = 1 - {}_nE_z - \bar{a}_{z\,\overline{n}\rceil}$

The life lengths of the statuses listed in the first table are increasing functions of both S and T. From the second table we see that, for a general status with remaining life length U, the present value of a pure survival benefit (life endowment or life annuity) is an increasing functions of U, whereas the present value of the pure death benefit is a decreasing function of U (assuming that the interest rate r is positive.)

Combining these observations and Theorem 1, we can infer the following (and many other things): If PQD(S,T), then pure survival benefits on any two statuses are positively dependent, pure death benefits on any two statuses are positively dependent, and any pure survival benefit and any death benefit are negatively dependent.

We can also draw conclusions about the bias introduced in equivalence premiums by erroneously adopting the independence hypothesis. For instance, if PQD(S,T) and we work under the independence hypothesis, then the present value of a survival benefit on the joint life will be underestimated, whereas the present value of a survival benefit on the last survivor will be overestimated. For the death benefit it is the other way around. Combining these things we may conclude e.g. that, for a death benefit on the joint life against level premium during joint survival, the equivalence premium will be overestimated.

D. A Markov chain model for two lives.
It is not easy to create a given form of dependence between the life lengths S and T by direct specification of their joint distribution. However, the process point of view, which is a powerful one, quite naturally allows us to express various ideas about dependencies between life lengths of a couple. A suitable framework is the Markov model sketched in Figure 7.4.

The following formulas are obvious:

$$
\begin{aligned}
p_{00}(s,t) &= e^{-\int_s^t \mu+\nu}\,, \\
p_{01}(s,t) &= \int_s^t e^{-\int_s^\tau \mu+\nu}\mu_\tau e^{-\int_\tau^t \nu'}\,d\tau\,, \\
p_{02}(s,t) &= \int_s^t e^{-\int_s^\tau \mu+\nu}\nu_\tau e^{-\int_\tau^t \mu'}\,d\tau\,.
\end{aligned}
$$

The joint survival function of S and T is

$$\begin{aligned}\mathbb{P}[S > s\,,\, T > t] &= \begin{cases} p_{00}(0,t) + p_{00}(0,s)p_{01}(s,t)\,, s \le t\,, \\ p_{00}(0,s) + p_{00}(0,t)p_{02}(t,s)\,, s > t\,, \end{cases} \\ &= \begin{cases} e^{-\int_0^t \mu+\nu} + \int_s^t e^{-\int_0^\tau \mu+\nu}\mu_\tau e^{-\int_\tau^t \nu'}\,, s \le t\,, \\ e^{-\int_0^s \mu+\nu} + \int_t^s e^{-\int_0^\tau \mu+\nu}\nu_\tau e^{-\int_\tau^s \mu'}\,, s > t\,. \end{cases} \end{aligned} \tag{7.97}$$

The marginal survival function of T is (put $s = 0$ in (7.97))

$$\begin{aligned}\mathbb{P}[T > t] &= p_{00}(0,t) + p_{01}(0,t) \\ &= e^{-\int_0^t \mu+\nu} + \int_0^t e^{-\int_0^\tau \mu+\nu}\mu_\tau e^{-\int_\tau^t \nu'}\,, t \ge 0\,. \end{aligned} \tag{7.98}$$

It is intuitively obvious that S and T are independent if $\mu'_\tau = \mu_\tau$ and $\nu'_\tau = \nu_\tau$ for all τ, and this will follow from Theorem 2 below. It is also intuitively clear that S and T will become dependent if we let the mortality rates depend on marital status. Let us see what happens if the mortality rate increases upon the loss of the spouse.

Theorem 2: *If $\mu'_\tau \ge \mu_\tau$ and $\nu'_\tau \ge \nu_\tau$ for all τ, then S and T are positively dependent in the sense RTI($S|T$) (hence AS(S,T) and PQD(S,T)).*

If $\mu'_\tau \le \mu_\tau$ and $\nu'_\tau \le \nu_\tau$ for all τ, then S and T are negatively dependent in the sense RTD($S|T$) (hence AS($-S,T$) and PQD($-S,T$)).

If $\mu'_\tau = \mu_\tau$ and $\nu'_\tau = \nu_\tau$ for all τ, then S and T are independent.

Proof: Consider first the case $s \le t$. From (7.97) and (7.98) we get

$$\begin{aligned}\mathbb{P}[S > s\,|\,T > t] &= \frac{e^{-\int_0^t \mu+\nu} + \int_s^t e^{-\int_0^\tau \mu+\nu}\,\mu_\tau\, e^{-\int_\tau^t \nu'}\,d\tau}{e^{-\int_0^t \mu+\nu} + \int_0^t e^{-\int_0^\tau \mu+\nu}\,\mu_\tau\, e^{-\int_\tau^t \nu'}\,d\tau} \\ &= 1 - \frac{\int_0^s e^{-\int_0^\tau \mu+\nu-\nu'}\,\mu_\tau\,d\tau}{e^{-\int_0^t \mu+\nu-\nu'} + \int_0^t e^{-\int_0^\tau \mu+\nu-\nu'}\,\mu_\tau\,d\tau}\,. \end{aligned}$$

Now we need only to study the denominator in the second term as a function of t. Its derivative is

$$e^{-\int_0^s \mu+\nu-\nu'}(\nu'_t - \nu_t)\,.$$

It follows that $\mathbb{P}[S > s\,|\,T > t]$ is an increasing function function of t if $\nu'_t \ge \nu_t$ and a decreasing function function of t if $\nu'_t \le \nu_t$.

The case $s > t$ is a bit more complicated. From (7.97) and (7.98) we get

$$\mathbb{P}[S > s\,|\,T > t] = \frac{e^{-\int_0^s \mu+\nu} + \int_t^s e^{-\int_0^\tau \mu+\nu}\,\nu_\tau\, e^{-\int_\tau^s \mu'}\,d\tau}{e^{-\int_0^t \mu+\nu} + \int_0^t e^{-\int_0^\tau \mu+\nu}\,\mu_\tau\, e^{-\int_\tau^t \nu'}\,d\tau}\,.$$

By the rule $d(u/v) = (v\,du - u\,dv)/v^2$, the sign of $\frac{\partial}{\partial t}\mathbb{P}[S > s\,|\,T > t]$ is the same as that of

$$\left(e^{-\int_0^t \mu+\nu} + \int_0^t e^{-\int_0^\tau \mu+\nu}\,\mu_\tau\, e^{-\int_\tau^t \nu'}\,d\tau\right)\left(-e^{-\int_0^t \mu+\nu}\,\nu_t\, e^{-\int_t^s \mu'}\right)$$

$$
\begin{aligned}
&-\left(e^{-\int_0^s \mu+\nu} + \int_t^s e^{-\int_0^\tau \mu+\nu}\,\nu_\tau\, e^{-\int_\tau^s \mu'}\,d\tau\right) \times \\
&\left(e^{-\int_0^t \mu+\nu}(-\mu_t-\nu_t) + e^{-\int_0^t \mu+\nu}\mu_t + \int_0^t e^{-\int_0^\tau \mu+\nu}\,\mu_\tau\, e^{-\int_\tau^t \nu'}\,d\tau\,(-\nu_t')\right).
\end{aligned}
$$

In this expression two terms cancel in the last parenthesis. Further, to get rid of some common factors, let us multiply with $e^{\int_0^s \mu+\nu}\, e^{\int_0^t \mu+\nu}$, which preserves the sign and turns the expression into

$$
\begin{aligned}
&-\left(1+\int_0^t e^{\int_\tau^t \mu+\nu-\nu'}\,\mu_\tau\,d\tau\right) e^{\int_t^s \mu-\mu'+\nu}\,\nu_t \\
&+\left(1+\int_t^s e^{\int_\tau^s \mu-\mu'+\nu}\nu_\tau\,d\tau\right)\left(\nu_t + \int_0^t e^{\int_\tau^t \mu+\nu-\nu'}\mu_\tau\,d\tau\,\nu_t'\right).
\end{aligned}
$$

Substituting

$$
\begin{aligned}
\int_t^s e^{\int_\tau^s \mu-\mu'+\nu}\nu_\tau\,d\tau &= \int_t^s e^{\int_\tau^s \mu-\mu'+\nu}(\mu_\tau-\mu_\tau'+\nu_\tau)\,d\tau \\
&\quad + \int_t^s e^{\int_\tau^s \mu-\mu'+\nu}(\mu_\tau'-\mu_\tau)\,d\tau \\
&= e^{\int_t^s \mu-\mu'+\nu} - 1 + \int_t^s e^{\int_\tau^s \mu-\mu'+\nu}(\mu_\tau'-\mu_\tau)\,d\tau
\end{aligned}
$$

and rearranging a bit, we arrive at

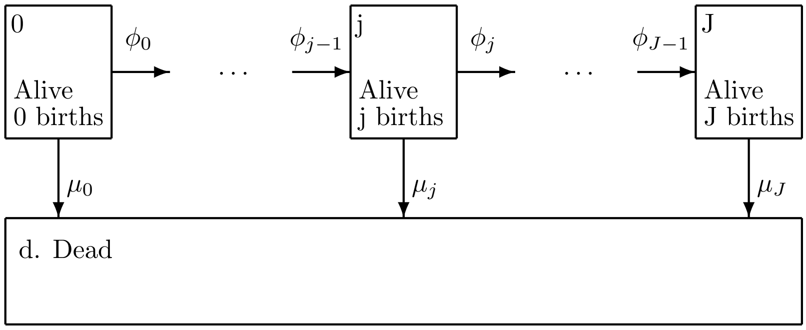
$$
\begin{aligned}
&\left(\nu_t + \nu_t' \int_0^t e^{\int_\tau^t \mu+\nu-\nu'}\mu_\tau\,d\tau\right)\int_t^s e^{\int_\tau^s \mu-\mu'+\nu}(\mu_\tau'-\mu_\tau)\,d\tau \\
&+\int_0^t e^{\int_\tau^s \mu+\nu-\nu'}\mu_\tau\,d\tau\,(\nu_t'-\nu_t).
\end{aligned}
$$

It follows that $\mathbb{P}[S>s\,|\,T>t]$ is an increasing function function of t if $\mu_t' \geq \mu_t$ and $\nu_t' \geq \nu_t$ and a decreasing function function of t if $\mu_t' \leq \mu_t$ and $\nu_t' \leq \nu_t$. □

7.10 Conditional Markov chains

A. Retrospective fertility analysis

In connection with a pension insurance scheme there is an additional benefit which is a sum insured payable to possible dependent children less than 18 years old at the time of death of the insured. In the technical basis we therefore need to make assumptions about births. We have to distinguish by sex, and in the following we consider female insured only. The Figure below shows a flowchart for possible life histories with death and births (at most J). To keep things simple, we assume that the insured enters the scheme in state 0 at age 0 and that the process is Markov: for a t year old who has given birth to j



children, the mortality rate $\mu_j(t)$ and the fertility rate $\phi_j(t)$ are functions of t and j only.

Assume now that the past history of births and death is observed only upon death of the insured, when the additional benefit to the possible dependents is due. Suppose that the statistical data comprise only those who are dead at the time of consideration and that for each of those there is a complete record of the times of possible births and of death. In these data the observed life history of a woman, who entered the scheme u years ago, is governed by a Markov process as described above, but with intensities

$$\mu_j^*(t) = \mu_j(t)\frac{1}{p_{jd}(t,u)}\,, \tag{7.99}$$

$$\phi_j^*(t) = \phi_j(t)\frac{p_{j+1,d}(t,u)}{p_{jd}(t,u)}\,. \tag{7.100}$$

We see that $\mu_j^*(t) \geq \mu_j(t)$, which is easy to explain (we are looking at the mortality, given death).

We are going to prove a more interesting result: If mortality increases with the number of births, that is,

$$\mu_j(t) \leq \mu_{j+1}(t)\,,\ j = 0, \ldots, J-1,\ t > 0\,, \tag{7.101}$$

then

$$\phi_j^*(t) \geq \phi_j(t)\,,\ j = 0, \ldots, J-1\,,\ t > 0\,. \tag{7.102}$$

We need to prove that $p_{j+1,d}(t,u) \geq p_{jd}(t,u)$, $j = 1, \ldots, J-1$. It is convenient to work with

$$p_j(t,u) = 1 - p_{jd}(t,u) = \sum_{k=j}^{J} p_{jk}(t,u)\,, \tag{7.103}$$

the probability that a t year old with j births will survive to age u, and to prove the hypothesis

$$H_j:\ p_k(t,u) \leq p_j(t,u)\,, \qquad k = j+1, \ldots, J, \tag{7.104}$$

for $j = 0, \ldots, J-1$. The proof goes by induction 'downwards', proving that H_{j+1} implies H_j. Thus assume H_{j+1} is true.

By direct reasoning (or an easy calculation) the mortality intensity at age u ($> t$) associated with the survival function (7.103) is

$$\mu_j(t,u) = \frac{\sum_{k \geq j} p_{jk}(t,u)\mu_k(u)}{\sum_{k \geq j} p_{jk}(t,u)}\,, \tag{7.105}$$

hence

$$p_j(t,u) = e^{-\int_t^u \mu_j(t,s)ds}\,. \tag{7.106}$$

Two more expressions for $p_j(t,u)$, both obvious, are

$$p_j(t,u) = \sum_{k \geq j} p_{jk}(t,\tau)\, p_k(\tau,u)\,, \tag{7.107}$$

$t \leq \tau \leq u$, and

$$p_j(t,u) = e^{-\int_t^u (\phi_j + \mu_j)} + \int_t^u e^{-\int_t^\tau (\phi_j+\mu_j)} \phi_j(\tau)\, p_{j+1}(\tau,u)\, d\tau\,. \tag{7.108}$$

By (7.101) and (7.105) we have

$$\mu_j(u) \leq \mu_{j+1}(t,u)\,, \tag{7.109}$$

hence

$$e^{-\int_t^u \mu_j} \geq p_{j+1}(t,u)\,.$$

Therefore, from (7.108) we get

$$\begin{aligned} p_j(t,u) \quad &\geq \quad e^{-\int_t^u \phi_j}\, p_{j+1}(t,u) \\ &\quad + \int_t^u e^{-\int_t^\tau \phi_j} \phi_j(\tau)\, p_{j+1}(t,\tau)\, p_{j+1}(\tau,u)\, d\tau\,. \end{aligned} \tag{7.110}$$

Focusing on the last two factors under the integral, use in succession (7.103), the induction hypothesis (7.104), and (7.107), to deduce

$$\begin{aligned} p_{j+1}(t,\tau)p_{j+1}(\tau,u) \quad &= \quad \sum_{k=j+1}^{J} p_{j+1,k}(t,\tau)p_{j+1}(\tau,u) \\ &\geq \quad \sum_{k=j+1}^{J} p_{j+1,k}(t,\tau)p_k(\tau,u) \\ &= \quad p_{j+1}(t,u)\,. \end{aligned}$$

Putting this into (7.110), we obtain

$$p_j(t,u) \geq \left(e^{-\int_t^u \phi_j} + \int_t^u e^{-\int_t^\tau \phi_j} \phi_j(t,\tau)\, d\tau\right) p_{j+1}(t,u) \;=\; p_{j+1}(t,u)\,.$$

It follows that H_j is true. Since H_{J-1} is obviously true, we are done. $\square$

Comment: The inequality (7.102) means that the fertility rates will be overestimated if one uses the estimators for the ϕ_j^* based on diseased participants in the scheme. If the inequalities (7.101) are reversed, then also the inequality (7.102) will be reversed, and the estimators the ϕ_j^* will underestimate the fertility. In particular it follows that, under the hypothesis of non-differential mortality, the fertility rates will be unbiasedly estimated from the selected material of diseased participants.

Chapter 8

Life history statistics

Think of the insurance company as a car:
At the steering wheel sits the chief executive
trying to keep the vehicle steady on the road.
In the front passenger seat sits the sales manager
pushing the speed pedal to the bottom.
At the rear sits the actuary peeping out the
back window and giving the directions.

8.1 Some principles of statistical inference

A. Learning from past experience. Any decision maker who believes there is some order and permanence in this world will use data from the past to make predictions about the future. Insurers, so it seems, adhere to such beliefs since they think systematic profit can be made from large scale engagement in risky contracts. Actuaries, who are the scribes and Pharisees of these convictions, invariably assume insurance claims are generated by random mechanisms that are simple enough to be described by well structured stochastic models and also sufficiently stable over time that statistics on past risk experience can be used to select and calibrate the models. The present chapter addresses the problem of estimating parameters in the life history models we have been working with up to now. Focus will be on maximum likelihood estimation of intensities in continuous time Markov chains. Before starting to fill the pages with all the particulars of that method and those models, let us state the problem of statistical inference in quite context-free terms.

B. The statistical model. Let θ represent the state of the nature or, rather, some aspect of it that we are interested in. The true state of the nature is unknown to us, but we assume – prior to empirical evidence – that it belongs to some given set ϑ of conceivable states. What *is* known to us is some observed data X that depends more or less on the state of the nature and thus carries some information as to which state in ϑ is the true one. Usually the data does not inform us fully about the state, and the uncertainty we are left with can be described by taking X to be random with a probability distribution $\mathbb{P}_\theta$ that depends on θ. The family $\{\mathbb{P}_\theta;\, \theta \in \vartheta\}$ constitutes our picture of the world, henceforth referred to as the *statistical model*.

C. Drawing inference from statistical data. Our problem is to infer something about the unknown state on the basis of the known data. More precisely, we seek a rule that to each possible outcome of the data delivers a statement about the true state of the nature. More precisely yet, we seek a function from the data space to some decision space, and this function should, to the extent possible given the uncertain nature of the data, deliver a correct statement. This problem may be formulated in various ways depending on the objective of the statistical inference. We shall list some possibilities.

D. Point estimation. If we just want to select a single point in ϑ as our best guess on the true state of the nature, then we are facing the problem of *point estimation* and we speak of the true state of the nature as the *estimand* (the quantity to be estimated). We must design a function $\hat{\theta}$, called *(point) estimator*, which to each outcome of X assigns a value $\hat{\theta}(X)$, called *(point) estimate*, in ϑ. The performance of an estimator is given by its *sampling distribution*, $\mathbb{P}_\theta[\hat{\theta}(X) \in B]$, $B \subset \vartheta$ (such that the probability is well defined), $\theta \in \vartheta$. A good estimator has a sampling distribution that is well concentrated (in some sense) around the estimand no matter what its true value is. For instance, if the space ϑ is equipped with some metric (measure of distance) Δ, we could try to find an estimator $\hat{\theta}$ such that $\int \Delta(\hat{\theta}(x), \theta)\, d\mathbb{P}_\theta(x)$ is small for all $\theta \in \vartheta$.

E. Region estimation. Due to the random nature of the data there will usually be not just one, but rather a whole set of points in ϑ that present themselves as reasonable estimates. We may then want to design a *region estimator* $\hat{\vartheta}$ which to each outcome x of the data X assigns a set $\hat{\vartheta}(x)$ of candidate true states in ϑ. If the region is an interval, we call it an *interval estimator*. The region estimator is said to have *confidence level* $1 - \varepsilon$ and is called a $1 - \varepsilon$ *confidence region* if it covers the estimand with a probability no less than $1 - \varepsilon$;

$$\mathbb{P}_\theta[\theta \in \hat{\vartheta}(X)] \geq 1 - \varepsilon\,, \qquad \forall \theta \in \vartheta. \tag{8.1}$$

The confidence we place in a region estimator is more worth the narrower the region is, of course. (For instance, the trivial region estimator $\hat{\vartheta} = \vartheta$ has confidence level 1, but is uninteresting.) Therefore, we might wish e.g. to minimize the size of the random set $\hat{\vartheta}(X)$ under a constraint on the confidence level, but this is usually too ambitious. In many situations an interval estimator arises naturally from knowledge about the sampling distribution of a point estimator and would appear as a point estimate equipped with error bounds.

F. Simultaneous confidence intervals. Let $\hat{\vartheta}$ be a $1-\varepsilon$ confidence region and denote by $\mathcal{G}$ the family of real-valued function defined on ϑ. For each $g \in \mathcal{G}$ define the random variables $\underline{g}(X) = \inf_{\theta \in \hat{\vartheta}(X)} g(\theta)$ and $\overline{g}(X) = \sup_{\theta \in \hat{\vartheta}(X)} g(\theta)$. Since the event $\left[\theta \in \hat{\vartheta}(X)\right]$ implies the event $\left[g(\theta) \in [\,\underline{g}(X), \overline{g}(X)\,]\,,\ \forall g \in \mathcal{G}\right]$, it follows from (8.1) that

$$\mathbb{P}_\theta\left[g(\theta) \in [\,\underline{g}(X), \overline{g}(X)\,]\,,\ \forall g \in \mathcal{G}\right] \geq 1-\varepsilon\,, \qquad \forall \theta \in \vartheta\,. \tag{8.2}$$

Thus, for each $g \in \mathcal{G}$ the interval $[\,\underline{g}(X), \overline{g}(X)\,]$ is a confidence interval for $g(\theta)$ with confidence level no less than $1-\varepsilon$. Since this confidence level holds for all functions $g \in \mathcal{G}$ together, the intervals $[\underline{g}(X), \overline{g}(X)]$ are called *simultaneous* $1-\varepsilon$ confidence intervals. A practical consequence is that 'data-snooping' is allowed; starting from a confidence region and the simultaneous confidence intervals derived from it, we can search for possible interesting effects (within the model), and still keep control of the probability of making any false statement.

G. Hypothesis testing. This term covers situations where we are interested in deciding whether or not the true state of the nature belongs to a given prespecified subset $\vartheta_0 \subset \vartheta$.

The problem can be set out in a clear-cut form in the context of scientific inquiry, where ϑ_0 represents a *hypothesis* whose truth we are particularly interested in and therefore want to test against empirical evidence. We need to design a decision rule δ, called a *test*, that delivers a decision whether to "accept" the hypothesis or to "reject" it. If the hypothesis is true, it should not be rejected, so one usually requires the test to satisfy

$$\mathbb{P}_\theta\left[\delta(X) = \text{reject}\right] \leq \varepsilon\,, \qquad \forall \theta \in \vartheta_0\,, \tag{8.3}$$

where ε is a chosen small number called the *level* of the test. Under this constraint one might seek to maximize $\mathbb{P}_\theta\left[\delta(X) = \text{reject}\right]$ for some alternative $\theta \notin \vartheta_0$. Only in certain simple statistical models and for certain simple ϑ_0 there exists a uniformly optimal level ε test that solves the maximization problem for all $\theta \notin \vartheta_0$.

If the hypothesis consists of just one point, $\vartheta_0 = \{\theta_0\}$, we call it *simple* and otherwise we call it *composite*. Obviously, from any level $1-\varepsilon$ confidence interval $\hat{\vartheta}$ we obtain a level ε test for the simple hypothesis $\{\theta_0\}$ by taking $\delta_{\theta_0}(X) =$ "reject" if and only if $\theta_0 \notin \hat{\vartheta}(X)$. Conversely, if for each $\theta \in \vartheta$, δ_θ is a level ε test for the simple hypothesis $\{\theta\}$, then we obtain a level $1-\varepsilon$ confidence interval by taking $\hat{\vartheta}(X) = \{\theta;\ \delta_\theta(X) = \text{“accept”}\}$. In view of Paragraph F above, it is also clear how we can test an arbitrary collection of hypotheses of the form $g(\theta) = g_0$ simultaneously with a probability no greater than ε of making any false rejection.

8.2 Maximum likelihood estimation

A. Maximum likelihood estimation. We assume henceforth that the model is *parametric*, which means that ϑ is of finite dimension r (say). Then its generic element is an r-dimensional parameter vector, $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_r)'$. Suppose also the data is finite-dimensional, $\mathbf{X} = (X_1, \ldots, X_p)$. Let $f(x_1, \ldots, x_p; \theta)$ denote its elementary probability function, which may be a probability (if the distribution is purely discrete) or a probability density (if the distribution is absolutely continuous) or a mixture of the two. To fix ideas, let us assume for the time being that $f(\,\cdot\,; \theta)$ is a density for all $\boldsymbol{\theta} \in \vartheta$ so that

$$\mathbb{P}_{\boldsymbol{\theta}}\left[X_i \in dx_i;\; i = 1, \ldots, p\right] = f(x_1, \ldots, x_p; \theta)\, dx_1 \cdots dx_n\,. \tag{8.4}$$

Here $X \in dx$ is short-hand for $x \leq X < x + dx$. The *likelihood function* L is obtained upon inserting the observed data in the elementary probability function,

$$L(\boldsymbol{\theta}; \mathbf{X}) = f(X_1, \ldots, X_p; \theta)\,,$$

and the *maximum likelihood estimator (MLE)* $\hat{\theta}$ is the value of $\boldsymbol{\theta}$ that maximizes the likelihood function;

$$\hat{\boldsymbol{\theta}}(\mathbf{X}) = \mathrm{argmax}_{\boldsymbol{\theta}} L(\boldsymbol{\theta}; \mathbf{X})\,.$$

In view of (8.4) the MLE is just the parameter value that best fits the actual observations in the sense of making them "as likely as possible".

The MLE holds a prominent place in the theory of statistical estimation, and within the statistical community there are congregations that take it as canonical and speak of the *maximum likelihood principle*. However, as was emphasized in the previous paragraph, any estimator has to judged by its sampling properties. One may construct situations where the MLE performs poorly. Fortunately, there exist theorems that settle this issue for broad classes of models, and we shall proceed with the ML method anticipating that it works well in the situations we are going to deal with here.

B. The case with a sample of independent observations. Let the data consist of random vectors $\mathbf{X}_1, \ldots, \mathbf{X}_n$ that are stochastically independent, and denote the elementary probability function of $\mathbf{X}_i$ by $f_i(\mathbf{x}_i; \theta)$, $\boldsymbol{\theta} \in \vartheta$. Then the likelihood function is

$$L_n(\boldsymbol{\theta}; \mathbf{X}_1, \ldots, \mathbf{X}_n) = \prod_{i=1}^{n} f_i(\mathbf{X}_i; \theta)\,, \tag{8.5}$$

and $f_i(\mathbf{X}_i; \theta)$ considered as a function of $\boldsymbol{\theta}$ is the 'individual likelihood function' of observation No. i. Obviously, the sample size n has a bearing on the amount of information in the data, hence the subscript n in the likelihood function and, later, in the MLE and certain related quantities.

Assume henceforth that ϑ is an open set and that the likelihood function is twice continuously differentiable. Maximizing the likelihood (8.5) amounts to setting its first order derivative with respect to each entry in $\boldsymbol{\theta}$ equal to 0 and solving. Since sums are easier to differentiate than products, we undertake to maximize the log likelihood function

$$\ln L_n(\boldsymbol{\theta}; \mathbf{X}_1, \ldots, \mathbf{X}_n) = \sum_{i=1}^{n} \ln f_i(\mathbf{X}_i; \theta)\,.$$

We differentiate it to obtain the so-called *score function*,

$$\frac{\partial}{\partial \boldsymbol{\theta}} \ln L_n(\boldsymbol{\theta}; \mathbf{X}_1, \ldots, \mathbf{X}_n) = \sum_{i=1}^{n} \frac{\partial}{\partial \boldsymbol{\theta}} \ln f_i(\mathbf{X}_i; \theta)\,, \tag{8.6}$$

and determine the MLE $\hat{\boldsymbol{\theta}}_n$ as the solution to the *ML equation*,

$$\sum_{i=1}^{n} \frac{\partial}{\partial \boldsymbol{\theta}} \ln f_i(\mathbf{X}_i; \hat{\boldsymbol{\theta}}_n) = \mathbf{0}\,. \tag{8.7}$$

The vector/matrix notation for derivatives used here may speak for itself, but we render a brief explanation. For instance, (8.7) is to be read as

$$\sum_{i=1}^{n} \frac{\partial}{\partial \theta_q} \ln f_i(\mathbf{X}_i; \hat{\boldsymbol{\theta}}_n) = 0\,, \qquad q = 1, \ldots, r\,.$$

Thus, the operator $\frac{\partial}{\partial \boldsymbol{\theta}}$ is the $r \times 1$ column operator vector whose q-th element is the operator $\frac{\partial}{\partial \theta_q}$. Likewise, $\frac{\partial}{\partial \boldsymbol{\theta}'}$ is the $1 \times r$ row operator vector whose q-th element is $\frac{\partial}{\partial \theta_q}$. We will also encounter $\frac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'}$, which is the symmetric $r \times r$ operator matrix with $\frac{\partial^2}{\partial \theta_p \partial \theta_q} = \frac{\partial}{\partial \theta_p} \frac{\partial}{\partial \theta_q} = \frac{\partial}{\partial \theta_q} \frac{\partial}{\partial \theta_p}$ in the p-th row and q-th column (and vice versa). Formally, $\frac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} = \frac{\partial}{\partial \boldsymbol{\theta}} \frac{\partial}{\partial \boldsymbol{\theta}'}$, where the latter is the (well defined) matrix product of an $r \times 1$ vector and a $1 \times r$ vector (of operators).

Since the individual score functions are the basic building blocks in the ML construction, we must expect that their statistical properties determine those of the MLE. It will turn out that their first two moments are crucial, so let us determine those for the i-th individual score function. Firstly, the expected value is identically zero:

$$\mathbb{E}\left[\frac{\partial}{\partial \boldsymbol{\theta}} \ln f_i(\mathbf{X}_i; \boldsymbol{\theta})\right] = \mathbf{0}\,. \tag{8.8}$$

This is seen by just spelling out the definition of the expected value:

$$\begin{aligned} \int \frac{1}{f_i(\mathbf{x}; \boldsymbol{\theta})} \frac{\partial}{\partial \boldsymbol{\theta}} f_i(\mathbf{x}; \theta) f_i(\mathbf{x}; \boldsymbol{\theta})\, d\mathbf{x} &= \int \frac{\partial}{\partial \boldsymbol{\theta}} f_i(\mathbf{x}; \boldsymbol{\theta})\, d\mathbf{x} = \frac{\partial}{\partial \boldsymbol{\theta}} \int f_i(\mathbf{x}; \boldsymbol{\theta})\, d\mathbf{x} \\ &= \frac{\partial}{\partial \boldsymbol{\theta}}(1) = \mathbf{0}\,. \end{aligned}$$

Secondly, the variance matrix, which is called the (individual) *information matrix* and here denoted by $\mathbf{J}_i(\boldsymbol{\theta})$, is

$$\mathbf{J}_i(\boldsymbol{\theta}) = \mathbb{V}_{\boldsymbol{\theta}}\left[\frac{\partial}{\partial\boldsymbol{\theta}}\ln f_i(\mathbf{X}_i;\boldsymbol{\theta})\right] \tag{8.9}$$

$$= \mathbb{E}_{\boldsymbol{\theta}}\left[\frac{\partial}{\partial\boldsymbol{\theta}}\ln f_i(\mathbf{X}_i;\boldsymbol{\theta})\frac{\partial}{\partial\boldsymbol{\theta}'}\ln f_i(\mathbf{X}_i;\boldsymbol{\theta})\right] \tag{8.10}$$

$$= \mathbb{E}_{\boldsymbol{\theta}}\left[-\frac{\partial^2}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\ln f_i(\mathbf{X}_i;\boldsymbol{\theta})\right]. \tag{8.11}$$

The expression (8.10) is just the definition of the variance (recall that the mean is 0). The expression (8.11) is obtained from the calculation

$$\begin{aligned}
\frac{\partial^2}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\ln f_i(\mathbf{X}_i;\boldsymbol{\theta}) &= \frac{\partial}{\partial\boldsymbol{\theta}}\left(\frac{1}{f_i(\mathbf{X}_i;\boldsymbol{\theta})}\frac{\partial}{\partial\boldsymbol{\theta}'}f_i(\mathbf{X}_i;\boldsymbol{\theta})\right)\\
&= -\frac{1}{f_{\boldsymbol{\theta}}^2(\mathbf{X}_i)}\frac{\partial}{\partial\boldsymbol{\theta}}f_i(\mathbf{X}_i;\boldsymbol{\theta})\frac{\partial}{\partial\boldsymbol{\theta}'}f_i(\mathbf{X}_i;\boldsymbol{\theta}) + \frac{1}{f_i(\mathbf{X}_i;\boldsymbol{\theta})}\frac{\partial^2}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}f_i(\mathbf{X}_i;\boldsymbol{\theta})\\
&= -\frac{\partial}{\partial\boldsymbol{\theta}}\ln f_i(\mathbf{X}_i;\boldsymbol{\theta})\frac{\partial}{\partial\boldsymbol{\theta}'}\ln f_i(\mathbf{X}_i;\boldsymbol{\theta}) + \frac{1}{f_i(\mathbf{X}_i;\boldsymbol{\theta})}\frac{\partial^2}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}f_i(\mathbf{X}_i;\boldsymbol{\theta})
\end{aligned}$$

and the fact that, similarly to (8.8),

$$\mathbb{E}_{\boldsymbol{\theta}}\left[\frac{1}{f_i(\mathbf{X}_i;\boldsymbol{\theta})}\frac{\partial^2}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}f_i(\mathbf{X}_i;\boldsymbol{\theta})\right] = \mathbf{0}.$$

We pause here to render some motivating intuitive considerations. Firstly, (8.8) adds a piece of motivation of the ML method. In view of (8.7), the MLE amounts to equating to zero something that has theoretical mean zero. Secondly, by virtue of (8.10), the variance matrix $\mathbf{J}_i(\boldsymbol{\theta})$ measures how strongly the individual score function depends on the outcome of observation No. i. A small variance means that the distribution of the observation is nearly the same for all parameter values and, therefore, the observation carries little information as to which parameter value is at work. The alternative expression (8.11) shows, what is intuitively appealing, that the amount of information depends on the sensitivity of the score function to changes in the parameter. Accordingly, the variance of the total score function in (8.6) can be taken as a measure of the total information contained in the data. It is called the (total) information matrix and denoted here by $\boldsymbol{\mathcal{J}}_n(\boldsymbol{\theta})$. Due to independence, the total information is the sum of the individual pieces of information:

$$\boldsymbol{\mathcal{J}}_n(\boldsymbol{\theta}) = \sum_{i=1}^{n}\mathbf{J}_i(\boldsymbol{\theta}). \tag{8.12}$$

We will now substantiate these preliminary ideas.

C. Asymptotic properties of the MLE. Provided the likelihood function satisfies certain regularity conditions and provided the total information increases to infinity,

$$\boldsymbol{\mathcal{J}}_n(\boldsymbol{\theta})^{-1} \to \mathbf{0}\,, \tag{8.13}$$

the MLE possesses the following two desirable asymptotic properties as $n \to \infty$. Firstly, its distribution tends to the normal distribution with mean equal to the estimand and with variance equal to the inverse of the total information matrix:

$$\hat{\boldsymbol{\theta}}_n \sim_{\text{as}} \mathrm{N}\left(\boldsymbol{\theta}, \boldsymbol{\mathcal{J}}_n(\boldsymbol{\theta})^{-1}\right)\,. \tag{8.14}$$

Thus, as the information increases the MLE becomes unbiased and its mean squared error decreases. This, ultimately, justifies the term information matrix. The precise meaning of (8.14) is

$$\boldsymbol{\mathcal{J}}_n(\boldsymbol{\theta})^{\frac{1}{2}}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}) \xrightarrow{\mathrm{L}} \mathrm{N}\left(\mathbf{0}, \mathbf{I}\right)\,, \tag{8.15}$$

where $\xrightarrow{\mathrm{L}}$ signifies convergence in distribution ("Law"). Secondly, the ML estimator is *consistent*, which means that it converges in probability to the estimand:

$$\hat{\boldsymbol{\theta}}_n \xrightarrow{\mathrm{P}} \boldsymbol{\theta}\,. \tag{8.16}$$

The regularity conditions required for (8.15) and (8.16) to hold true are rather technical, but essentially they just assume that the log likelihood function possesses derivatives up to order three and that these derivatives can be bounded uniformly by random variables that are 'sufficiently integrable'. For a precise statement of these conditions and a rigorous proof in the simple case with a scalar parameter and i.i.d. observations, see e.g. Serfling [45]. Aiming at more general situations, we shall be content here to just indicate how (8.16) and (8.15) are obtained via a linear approximation of the score function.

Upon expanding the expression on the left in the ML equation (8.7) in a first order Taylor series around the estimand $\boldsymbol{\theta}$ and simply ignoring the remainder term, the relationship becomes

$$\sum_{i=1}^{n} \frac{\partial}{\partial \boldsymbol{\theta}} \ln f_i(\mathbf{X}_i; \boldsymbol{\theta}) + \sum_{i=1}^{n} \frac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \ln f_i(\mathbf{X}_i; \boldsymbol{\theta})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) = \mathbf{0}\,.$$

Solving with respect to $\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}$ and premultiplying with the square root of the total information, gives

$$\boldsymbol{\mathcal{J}}_n(\boldsymbol{\theta})^{\frac{1}{2}}\left(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}\right) = \mathbf{U}_n^{-1}\,\mathbf{v}_n\,, \tag{8.17}$$

where

$$\mathbf{U}_n = \boldsymbol{\mathcal{J}}_n(\boldsymbol{\theta})^{-\frac{1}{2}} \left(\sum_{i=1}^{n} -\frac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \ln f_i(\mathbf{X}_i; \boldsymbol{\theta}) \right) \boldsymbol{\mathcal{J}}_n(\boldsymbol{\theta})^{-\frac{1}{2}}\,, \tag{8.18}$$

$$\mathbf{v}_n = \boldsymbol{\mathcal{J}}_n(\boldsymbol{\theta})^{-\frac{1}{2}} \sum_{i=1}^{n} \frac{\partial}{\partial \boldsymbol{\theta}} \ln f_i(\mathbf{X}_i; \boldsymbol{\theta}) \,. \tag{8.19}$$

Loosely speaking, in one-dimensional language, the factor $\mathbf{U}_n$ is essentially a sum of independent random variables divided by the sum of their (strictly positive) expected values, and therefore typically converges in probability to one:

$$\mathbf{U}_n \xrightarrow{\mathrm{P}} \mathbf{I} \,. \tag{8.20}$$

The vector $\mathbf{v}_n$ is essentially a standardized sum of independent zero mean random variables, and therefore its distribution typically converges to the standard normal:

$$\mathbf{v}_n \xrightarrow{\mathrm{L}} \mathrm{N}\left(\mathbf{0}, \mathbf{I}\right) \,. \tag{8.21}$$

Combining these two conjectures, assuming they are true, we obtain (8.15). The consistency property (8.16) then follows from (8.13).

These matters become fully transparent in the special case where the observations are i.i.d. so that $f_i = f$ and $\mathbf{J}_i(\boldsymbol{\theta}) = \mathbf{J}(\boldsymbol{\theta})$, independent of i, and $\boldsymbol{\mathcal{J}}_n(\boldsymbol{\theta}) = n\mathbf{J}(\boldsymbol{\theta})$. Then (8.18) and (8.19) reduce to

$$\mathbf{U}_n = \mathbf{J}(\boldsymbol{\theta})^{-\frac{1}{2}} \left(\frac{1}{n} \sum_{i=1}^{n} - \frac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \ln f(\mathbf{X}_i; \boldsymbol{\theta}) \right) \mathbf{J}(\boldsymbol{\theta})^{-\frac{1}{2}} \,,$$

$$\mathbf{v}_n = \mathbf{J}(\boldsymbol{\theta})^{-\frac{1}{2}} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\partial}{\partial \boldsymbol{\theta}} \ln f(\mathbf{X}_i; \boldsymbol{\theta}) \,,$$

and (8.20) and (8.21) follow from of the standard versions of the law of large numbers and the central limit theorem, respectively. In this case (8.13) is trivially fulfilled.

In the scalar case, $r = 1$, it is usually easy to verify (8.20) and (8.21) directly. We will not pursue this any further since the purpose of the whole exercise was just to motivate the main results, and we have been ignoring the remainder term anyway.

In some simple situations one obtains explicit expressions for the MLE, and it may then be possible to prove (8.20) and (8.21) directly. One such case will be discussed in detail in the Section 8.4.

8.3 Mortality investigations

A. Incompletely observed life lengths. In the general introduction to ML estimation we deliberately abstained from making the assumption that the observations are i.i.d., which would have made things simpler. The i.i.d. assumption is appropriate in situations where the observations are outcomes of independent replicates of some designed experiment (e.g. repeated measurements of one and the same quantity) or random selections from a large population. Now, insurance is not an experimental discipline and, typically, observations gathered in connection with insurance operations are not made under identical circumstances. This feature stands out clearly in the context of mortality studies.

Let $T_1, \ldots, T_n$ be i.i.d. positive random variables representing the life lengths of n individuals sampled from some population. Their common distribution function and density are suitably expressed in terms of the mortality intensity, $\mu(t; \boldsymbol{\theta})$, which is head and tail of any mortality law:

$$F(t; \boldsymbol{\theta}) = 1 - e^{-\int_0^t \mu(s; \boldsymbol{\theta})\, ds}, \qquad f(t; \boldsymbol{\theta}) = e^{-\int_0^t \mu(s; \boldsymbol{\theta})\, ds} \mu(t; \boldsymbol{\theta}) . \tag{8.22}$$

If the i.i.d. total life lengths should be our observations, then the n individuals would have to be selected in a random manner at their times of birth and observed continually until death. Such an observational scheme would be rare in real life (and death) mortality investigations; the estimates are needed today, not only sooner or later when the last survivor has passed away. In practice an individual will be observed only over a certain cut of its life history, and this cut will vary among individuals. A case in point is the typical mortality study conducted by a life insurance company. The data consists of the policy records for all customers previously or currently insured under a certain insurance scheme. Let n denote the number of such records. Customer No. i entered the portfolio at age x_i, at the date when the policy was issued, and has thereafter been continually observed for z_i years, until the date when the contract terminated or the date when the data were extracted, whichever occurred first. Such a partially observed life history is said to be *left-censored* at age x_i and *right-censored* at age $y_i = x_i + z_i$. Moreover, it is known that the customer was alive at age x_i. Thus, what is observed is, not the total life length T_i, but rather the truncated life length $T_i \wedge y_i$, given that $T_i > x_i$. Its cumulative distribution function is

$$F_i(t; \boldsymbol{\theta}) = \mathbb{P}_{\boldsymbol{\theta}}[T_i \wedge y_i \le t \mid T_i > x_i] = \begin{cases} 1 - e^{-\int_{x_i}^t \mu(s; \boldsymbol{\theta})\, ds}, & x_i \le t < y_i , \\ 1, & t \ge y_i . \end{cases} \tag{8.23}$$

The elememtary probability function of this distribution is a mixture of a density on the interval (x_i, y_i) and a point proability at y_i:

$$f_i(t; \boldsymbol{\theta}) = \begin{cases} e^{-\int_{x_i}^t \mu(s; \boldsymbol{\theta})\, ds} \mu(t; \boldsymbol{\theta}), & x_i < t < y_i , \\ e^{-\int_{x_i}^{y_i} \mu(s; \boldsymbol{\theta})\, ds}, & t = y_i . \end{cases} \tag{8.24}$$

The individual likelihood can be gathered in the single expression

$$f_i(T_i \wedge y_i; \boldsymbol{\theta}) = \mu(T_i; \boldsymbol{\theta})^{N_i} e^{-\int_{x_i}^{T_i \wedge y_i} \mu(s; \boldsymbol{\theta})\, ds} ,$$

where

$$N_i = 1[x_i < T_i \le y_i]$$

is the 'number of deaths' of individual No. i during the observation period. The individual score function is

$$\frac{\partial}{\partial \boldsymbol{\theta}} \ln f_i(T_i \wedge y_i; \boldsymbol{\theta}) = N_i \frac{\partial}{\partial \boldsymbol{\theta}} \ln \mu(T_i; \boldsymbol{\theta}) - \int_{x_i}^{T_i \wedge y_i} \frac{\partial}{\partial \boldsymbol{\theta}} \mu(s; \boldsymbol{\theta}) \, ds \,, \tag{8.25}$$

and the ML equation becomes

$$\sum_{i; N_i = 1} \frac{1}{\mu(T_i; \boldsymbol{\theta})} \frac{\partial}{\partial \boldsymbol{\theta}} \mu(T_i; \boldsymbol{\theta}) - \sum_{i=1}^{n} \int_{x_i}^{T_i \wedge y_i} \frac{\partial}{\partial \boldsymbol{\theta}} \mu(s; \boldsymbol{\theta}) \, ds = \mathbf{0} \,. \tag{8.26}$$

We will show that the individual information matrix is given by

$$\mathbf{J}_i(\boldsymbol{\theta}) = \int_{x_i}^{y_i} \frac{1}{\mu(t; \boldsymbol{\theta})} \frac{\partial}{\partial \boldsymbol{\theta}} \mu(t; \boldsymbol{\theta}) \frac{\partial}{\partial \boldsymbol{\theta}'} \mu(t; \boldsymbol{\theta}) \, e^{-\int_{x_i}^{t} \mu(s; \boldsymbol{\theta})(s) \, ds} \, dt \,. \tag{8.27}$$

The pedestrian route to this result is to compute the expected value (8.10) or (8.11) (the latter seems more convenient) in the conditional distribution (8.23). An alternative and much simpler approach makes use of the indicator and counting processes associated with the life length T_i. Thus, introduce

$$I_i(t) = 1[T_i > t] \,, \quad N_i(t) = 1[T_i \leq t] \,,$$

and start afresh from (8.25) recast as

$$\begin{aligned} \frac{\partial}{\partial \boldsymbol{\theta}} \ln f_i(T_i \wedge y_i; \boldsymbol{\theta}) &= \int_{x_i}^{y_i} \left(\frac{\partial}{\partial \boldsymbol{\theta}} \ln \mu(t; \boldsymbol{\theta}) dN_i(t) - \frac{\partial}{\partial \boldsymbol{\theta}} \mu(t; \boldsymbol{\theta}) I_i(t) \, dt \right) \\ &= \int_{x_i}^{y_i} \frac{\partial}{\partial \boldsymbol{\theta}} \ln \mu(t; \boldsymbol{\theta}) \left(dN_i(t) - \mu(t; \boldsymbol{\theta}) I_i(t) \, dt \right) \,. \end{aligned} \tag{8.28}$$

We rediscover (8.8) by noting that

$$\mathbb{E}_{\boldsymbol{\theta}}[I_i(t) \,|\, I_i(x_i) = 1] = e^{-\int_{x_i}^{t} \mu(s; \boldsymbol{\theta}) \, ds} \,, \tag{8.29}$$

$$\mathbb{E}_{\boldsymbol{\theta}}[dN_i(t) \,|\, I_i(x_i) = 1] = e^{-\int_{x_i}^{t} \mu(s; \boldsymbol{\theta}) \, ds} \mu(t; \boldsymbol{\theta}) \, dt \,. \tag{8.30}$$

Differentiating (8.28) gives

$$\begin{aligned} \frac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \ln f_i(T_i \wedge y_i; \boldsymbol{\theta}) &= \int_{x_i}^{y_i} \frac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \ln \mu(t; \boldsymbol{\theta}) \left(dN_i(t) - \mu(t; \boldsymbol{\theta}) I_i(t) \, dt \right) \\ &\quad - \int_{x_i}^{y_i} \frac{\partial}{\partial \boldsymbol{\theta}} \ln \mu(t; \boldsymbol{\theta}) \frac{\partial}{\partial \boldsymbol{\theta}'} \mu(t; \boldsymbol{\theta}) I_i(t) \, dt \,. \end{aligned}$$

Forming conditional expectation, given $I_i(x_i) = 1$, and using (8.29) and (8.30) again, we arrive at (8.27).

The simplicity of this method adds another notch to the merits of the process point of view. It was not really needed so far and, strictly speaking, we might manage without it also in the following, but it will simplify matters greatly as we turn to more advanced problems.

B. The Gompertz-Makeham mortality law. We will work out the details for the case where the intensity is of the form

$$\mu(t,\boldsymbol{\theta}) = \alpha + \beta e^{\gamma t}\,. \tag{8.31}$$

The parameter is now three-dimensional,

$$\boldsymbol{\theta} = \begin{pmatrix} \alpha \\ \beta \\ \gamma \end{pmatrix},$$

with all entries non-negative.

In the ML equations (8.26) we need the derivatives

$$\frac{\partial}{\partial\alpha}\mu(t,\boldsymbol{\theta}) = 1\,, \tag{8.32}$$

$$\frac{\partial}{\partial\beta}\mu(t,\boldsymbol{\theta}) = e^{\gamma t}\,, \tag{8.33}$$

$$\frac{\partial}{\partial\gamma}\mu(t,\boldsymbol{\theta}) = \beta\, e^{\gamma t}\, t\,, \tag{8.34}$$

and their integrals,

$$\begin{aligned}
\int_x^y \frac{\partial}{\partial\alpha}\mu(t,\boldsymbol{\theta})\,dt &= y - x\,,\\
\int_x^y \frac{\partial}{\partial\beta}\mu(t,\boldsymbol{\theta})\,dt &= \frac{e^{\gamma y} - e^{\gamma x}}{\gamma}\,,\\
\int_x^y \frac{\partial}{\partial\gamma}\mu(t,\boldsymbol{\theta})\,dt &= \beta\left(\frac{e^{\gamma y}y - e^{\gamma x}x}{\gamma} - \frac{e^{\gamma y} - e^{\gamma x}}{\gamma^2}\right).
\end{aligned}$$

The MLE equations become

$$\begin{aligned}
&\sum_{i;\,N_i=1} \frac{1}{\alpha + \beta e^{\gamma T_i}} - \sum_i ((T_i \wedge y_i) - x_i) = 0\,,\\
&\sum_{i;\,N_i=1} \frac{e^{\gamma T_i}}{\alpha + \beta e^{\gamma T_i}} - \sum_i \frac{e^{\gamma (T_i\wedge y_i)} - e^{\gamma x_i}}{\gamma} = 0\,,\\
&\sum_{i;\,N_i=1} \frac{e^{\gamma T_i}T_i}{\alpha + \beta e^{\gamma T_i}} - \sum_i \left(\frac{e^{\gamma (T_i\wedge y_i)}(T_i \wedge y_i) - e^{\gamma x_i}x_i}{\gamma} - \frac{e^{\gamma (T_i\wedge y_i)} - e^{\gamma x_i}}{\gamma^2}\right) = 0\,.
\end{aligned}$$

These equations are highly non-linear and allow no explicit solution. One must resort to numerical methods to compute the MLE estimate $\hat{\boldsymbol{\theta}} = (\hat{\alpha}, \hat{\beta}, \hat{\gamma})'$. A Newton-Raphson procedure can easily be arranged.

In the individual information matrices (8.27) we need the inverse of the intensity (8.31), the cross products of the derivatives in (8.32) – (8.34),

$$\frac{\partial}{\partial\theta}\mu(t,\boldsymbol{\theta})\frac{\partial}{\partial\theta'}\mu(t,\boldsymbol{\theta}) = \begin{pmatrix} 1 & e^{\gamma t} & \beta e^{\gamma t}t \\ \cdot & e^{2\gamma t} & \beta e^{2\gamma t}t \\ \cdot & \cdot & \beta^2 e^{2\gamma t}t^2 \end{pmatrix}, \tag{8.35}$$

(symmetric) and the survival probabilities

$$e^{-\int_{x_i}^{t}(\alpha+\beta e^{\gamma\, s})\, ds} = e^{-\alpha(t-x_i)-\beta(e^{\gamma t}-e^{\gamma x_i})/\gamma}\,. \tag{8.36}$$

The total information matrix is

$$\boldsymbol{\mathcal{J}}_n(\boldsymbol{\theta}) = \int \frac{e^{-\alpha t-\beta e^{\gamma t}/\gamma}}{\alpha+\beta e^{\gamma t}} \begin{pmatrix} 1 & e^{\gamma t} & \beta e^{\gamma t} t \\ e^{\gamma t} & e^{2\gamma t} & \beta e^{2\gamma t} t \\ \beta e^{\gamma t} t & \beta e^{2\gamma t} t & \beta^2 e^{2\gamma t} t^2 \end{pmatrix} \sum_{i; x_i<t\le y_i} e^{\alpha x_i+\beta e^{\gamma x_i}/\gamma} dt\,.$$

The integral ranges effectively over the interval from $\min_i x_i$ to $\max_i y_i$ where there may be participants in the study. It is best computed by solving the obvious differential equations for the integral.

The censoring scheme affects only the factor $\sum_{i;x_i<t\le y_i} e^{\alpha x_i+\beta e^{\gamma x_i}/\gamma} dt$ in the integrand. It is seen that the information increases with the number of observations

8.4 The exponential survival function

A. Exponentially distributed life lengths. We will investigate in some detail the simple case where the individual life lengths T_i, $i = 1, \ldots, n$ are i.i.d. and exponentially distributed. Thus, assume the intensity μ is independent of age so that (8.22) reduces to

$$F(t;\mu) = 1 - e^{-\mu\, t}\,, \qquad f(t;\mu) = e^{-\mu\, t}\mu\,. \tag{8.37}$$

Since the remaining life length of an x year old is distributed as the total life length of a newborn, we can as well assume that all x_i are 0 and that our observations are the truncated total life lengths

$$W_i = T_i \wedge z_i = \int_0^{z_i} I_i(t)\, dt\,,$$

$i = 1, \ldots, n$. The number of deaths (0 or 1) of individual No. i during its period of observation is

$$N_i = 1[T_i \le z_i] = 1 - I_i(z_i) = \int_0^{z_i} dN_i(t)\, dt\,.$$

Now the unknown parameter is just the scalar μ. The likelihood of the observations becomes

$$L = \Pi_{i=1}^n \mu^{N_i} e^{-\mu W_i} = e^{\ln\mu\, \sum_{i=1}^n N_i - \mu \sum_{i=1}^n W_i}\,. \tag{8.38}$$

The ML equations (8.26) reduce to the single equation,

$$\frac{1}{\mu}\sum_{i=1}^n N_i - \sum_{i=1}^n W_i = 0\,,$$

which has the explicit solution

$$\hat{\mu}_n = \frac{\sum_{i=1}^n N_i}{\sum_{i=1}^n W_i} \,. \tag{8.39}$$

Thus, the MLE is the so-called *occurrence-exposure rate* which is the ratio of number of deaths divided by the total time spent at risk of death in the study. It is the straightforward empirical counterpart of the mortality intensity, which is the expected number of deaths per time unit and per survivor. The individual information matrix (8.27) reduces to the scalar

$$\int_0^{z_i} \frac{1}{\mu} e^{-\mu\, t}\, dt = \frac{1 - e^{\mu\, z_i}}{\mu^2} \,, \tag{8.40}$$

and the total information is

$$\mathcal{J}_n(\mu) = \frac{\sum_{i=1}^n (1 - e^{-\mu\, z_i})}{\mu^2} \,.$$

The asymptotic results (8.14) and (8.16) now read

$$\hat{\mu}_n \sim_{\text{as}} \mathrm{N}\left(\mu, \frac{\mu^2}{\sum_{i=1}^n (1 - e^{-\mu\, z_i})}\right) \,, \tag{8.41}$$

$$\hat{\mu}_n \xrightarrow{\mathrm{P}} \mu \,. \tag{8.42}$$

In the present situation we do not need to rely on these asymptotic results (which we have not proved rigorously). They can be worked out from first principles since the MLE is explicitly given by (8.39). To this end we need the following formulas:

$$\begin{aligned}
\mathbb{E}[N_i^k] &= F(z_i)\,,\ k = 1, 2, \ldots\,, & (8.43)\\
\mathbb{E}[W_i^k] &= k \int_0^{z_i} t^{k-1}(1 - F(t))\, dt\,,\ k = 1, 2, \ldots\,, & (8.44)\\
\mathbb{E}[N_i W_i] &= \mathbb{E}[W_i] - z_i(1 - F(z_i))\,. & (8.45)
\end{aligned}$$

Formula (8.43) is obvious since N_i only assumes values 0 and 1. To prove (8.44), use the chain rule and the fact that $I_i(\tau)I_i(t) = I_i(t)$ for $\tau \le t$ to write

$$\left(\int_0^{z_i} I_i(t)\, dt\right)^k = \int_0^{z_i} k \left(\int_0^t I_i(\tau)\, d\tau\right)^{k-1} I_i(t)\, dt = k \int_0^{z_i} t^{k-1} I_i(t)\, dt \,,$$

and form expectation. To prove (8.45), write

$$N_i W_i = (1 - I_i(z_i))W_i = W_i - I_i(z_i) \int_0^{z_i} I_i(t)\, dt = W_i - I_i(z_i) z_i\, dt$$

and form expectation. Using these results together with (8.37) one easily calculates

$$\mathbb{E}[N_i] = 1 - e^{-\mu z_i}\,, \tag{8.46}$$
$$\mathbb{V}[N_i] = e^{-\mu z_i}(1 - e^{-\mu z_i})\,, \tag{8.47}$$
$$\mathbb{E}[W_i] = \frac{1 - e^{-\mu z_i}}{\mu}\,, \tag{8.48}$$
$$\mathbb{V}[W_i] = \frac{1 - 2\mu z_i e^{-\mu z_i} - e^{-2\mu z_i}}{\mu^2}\,, \tag{8.49}$$
$$\mathbb{E}[N_i - \mu W_i] = 0\,, \tag{8.50}$$
$$\mathbb{V}[N_i - \mu W_i] = 1 - e^{-\mu z_i}\,. \tag{8.51}$$

B. The special case with no censoring. Suppose now that the n lives are completely observed without censoring, that is, $z_i = \infty$ and $W_i = T_i$, $i = 1, \ldots, n$. Then all N_i are 1, $\sum_{i=1}^n N_i = n$, and $\sum_{i=1}^n W_i = \sum_{i=1}^n T_i$. In this simple situation it is easy to investigate the small sample properties of the MLE. The sum of the life lengths, $W = \sum_{i=1}^n T_i$, is now a sufficient statistic. It has a gamma distribution with shape parameter n and scale parameter $\nu = 1/\mu$, whose density is

$$\frac{\mu^n}{\Gamma(n)} w^{n-1} e^{-\mu w}\ ,\ w > 0\,.$$

One finds (perform the easy calculations) for $k > -n$ that

$$\mathbb{E}[W^k] = \frac{\Gamma(n+k)}{\Gamma(n)\mu^k}\,,$$

hence

$$\mathbb{E}[\hat{\mu}] = \frac{n\mu}{n-1}\ ,\ n > 1\,,$$

and

$$\mathbb{V}[\hat{\mu}] = \frac{n^2\mu^2}{(n-1)^2(n-2)}\ ,\ n > 2\,.$$

The estimator $\hat{\mu}$ is biased and, on the average, overestimates μ by $\mu/(n-1)$, which is negligible for large n. An unbiased estimator of μ is $\check{\mu} = (n-1)/W$. Its variance is $\mu^2/(n-2)$.

C. Asymptotic results by uniform censoring. Suppose all z_i are equal to z, say. Writing

$$\hat{\mu} = \frac{\frac{1}{n}\sum_{i=1}^n N_i}{\frac{1}{n}\sum_{i=1}^n W_i}\,,$$

and noting that $\mathbb{E}[N_i] = \mu\mathbb{E}T_i$ by (8.46) and (8.48), it follows by the weak law of large numbers that the estimator is consistent,

$$\hat{\mu} \xrightarrow{\mathrm{P}} \mu\,.$$

To investigate its asymptotic distribution, look at

$$\sqrt{n}(\hat{\mu} - \mu) = \frac{\frac{1}{\sqrt{n}} \sum_{i=1}^{n} (N_i - \mu W_i)}{\frac{1}{n} \sum_{i=1}^{n} W_i} .$$

The denominator of this ratio converges a.s. to $\mathbb{E}[W_i]$ given by (8.48). By the central limit theorem, the limiting distribution of the numerator is normal with mean 0 (recall (8.50)) and variance given by (8.51). It follows that

$$\hat{\mu} \sim_{\text{as}} \mathrm{N}\left(\mu, \frac{\mu^2}{n(1 - e^{-\mu z})}\right) . \tag{8.52}$$

D. Asymptotic results by fairly general censoring. Consider now the general situation in Paragraph A with censoring varying among the individuals. A bit more effort must now be put into the study of the asymptotic properties of the MLE. It turns out that a sufficient condition for consistency and asymptotic normality is that the expected exposure grows to infinity in the sense that

$$\sum_{i=1}^{n} \mathbb{E}[W_i] \to \infty ,$$

which by (8.48) is equivalent to

$$\sum_{i=1}^{n} (1 - e^{-\mu z_i}) \to \infty , \tag{8.53}$$

that is, the expected number of deaths grows to infinity. Thus assume that (8.53) is satisfied. In the following the relationships (8.46) – (8.51) will be used frequently without explicit mentioning.

First, to prove consistency, use (8.48) and (8.51) to write

$$\begin{aligned} \hat{\mu} - \mu &= \frac{\sum_{i=1}^{n} (N_i - \mu W_i)}{\sum_{i=1}^{n} W_i} . \\ &= \frac{\sum_{i=1}^{n} (N_i - \mu W_i)}{\sum_{i=1}^{n} \mathbb{V}[N_i - \mu W_i]} \left(\frac{\sum_{i=1}^{n} W_i}{\mu \sum_{i=1}^{n} \mathbb{E}[W_i]} \right)^{-1} . \end{aligned} \tag{8.54}$$

The first factor in (8.54) has expected value 0 and variance

$$\frac{1}{\sum_{i=1}^{n} \mathbb{V}[N_i - \mu W_i]} = \frac{1}{\sum_{i=1}^{n} (1 - e^{-\mu z_i})} ,$$

which tends to 0 as n increases. Therefore, this factor tends to 0 in probability. The second factor in (8.54) is the inverse of $\sum_{i=1}^{n} W_i / \mu \sum_{i=1}^{n} \mathbb{E}[W_i]$, which has expected value $1/\mu$ and variance equal to $1/\mu^2$ times

$$\begin{aligned} \frac{\sum_{i=1}^{n} \mathbb{V}[W_i]}{\left(\sum_{i=1}^{n} (1 - e^{-\mu z_i})\right)^2} &= \frac{\sum_{i=1}^{n} (1 - 2\mu z_i e^{-\mu z_i} - e^{-2\mu z_i})}{\left(\sum_{i=1}^{n} (1 - e^{-\mu z_i})\right)^2} \\ &= \frac{\sum_{i=1}^{n} a(\mu z_i)(1 - e^{-\mu z_i})}{\sum_{i=1}^{n} (1 - e^{-\mu z_i})} \frac{1}{\sum_{i=1}^{n} (1 - e^{-\mu z_i})} , \end{aligned} \tag{8.55}$$

where a is defined as

$$a(t) = \frac{1 - 2te^{-t} - e^{-2t}}{1 - e^{-t}}\ ,\ t \geq 0.$$

The function a is bounded since it is continuous and tends to 0 as $t \searrow 0$ (use l'Hospital's rule) and to 1 as $t \nearrow \infty$. The first factor in (8.55) is bounded since it is a weighted average of values of a, and the second factor tends to 0 by assumption. It follows that the expression in (8.55) tends to 0 and, consequently, that the second factor in (8.54) converges in probability to μ. It can be concluded that the expression in (8.54) converges in probability to 0, so that $\hat{\mu}$ is weakly consistent;

$$\hat{\mu} \xrightarrow{\mathrm{P}} \mu\,.$$

Next, to prove asymptotic normality, look at

$$\sqrt{\sum_{i=1}^{n}(1 - e^{-\mu z_i})}(\hat{\mu} - \mu) = \frac{\sum_{i=1}^{n}(N_i - \mu W_i)}{\sqrt{\sum_{i=1}^{n}(1 - e^{-\mu z_i})}}\left(\frac{\sum_{i=1}^{n} W_i}{\mu \sum_{i=1}^{n} \mathbb{E}[W_i]}\right)^{-1}.$$

In the presence of (8.53) the first factor on the right converges in distribution to a standard normal variate. (Verify that the Lindeberg condition is satisfied, see Appendix D.)

The second factor has just been proved to converge in probability to μ. It follows that

$$\hat{\mu} \sim \mathrm{as}\left(\mu, \frac{\mu^2}{\sum_{i=1}^{n}(1 - e^{-\mu z_i})}\right). \tag{8.56}$$

E. Random censoring. In Paragraph C the censoring time was assumed to be the same for all individuals. Thereby the pairs (N_i, W_i) became stochastic replicates, and we could invoke simple asymptotic theory for i.i.d. variates to prove strong consistency and asymptotic normality of MLE-s. In Paragraph C the censoring was allowed to vary among the individuals, but it turned out that the asymptotic results essentially remained true, although only weak consistency could be achieved. All we required was (8.53), which says that the censoring must not turn too severe so that information deteriorates in the end: there must be a certain stability in the censoring pattern so that individuals with sufficient exposure time enter the study sufficiently frequently in the long run. One way of securing such stability is to regard the censoring times as outcomes of i.i.d. random variables. Such an assumption seems particularly apt in a non-experimental context like insurance. The censoring is not subject to planning, and the censoring times are just as random in their nature as anything else observed about the individuals.

Thus, we henceforth work with an augmented model where the assumptions in Paragraph D constitute the conditional model for given censoring times $Z_i = z_i$, $i = 1, 2, \ldots$, and the Z_i are independent selections from some distribution function H with (generalized) density h independent of μ. This way the triplets

(N_i, T_i, Z_i), $i = 1, 2, \ldots$, become stochastic replicates, and the i.i.d. situation is restored with all its powers.

The likelihood of the observations now becomes

$$L = \Pi_{i=1}^{n} \mu^{N_i} e^{-\mu W_i} h(Z_i) = e^{\ln \mu \sum_{i=1}^{n} N_i - \mu \sum_{i=1}^{n} W_i} \Pi_{i=1}^{n} h(Z_i) \,. \tag{8.57}$$

Maximizing (8.57) with respect to μ is equivalent to maximizing the likelihood (8.38) in the conditional model for fixed censoring, hence the MLE remains the same as before. Its distribution is affected by the structure now added to the model, however. It is easy to prove that the results in Paragraph D carry over to the present case, only that the expression $1 - e^{-\mu z}$ is everywhere to be replaced by $1 - \mathbb{E}\left[e^{-\mu Z}\right]$, where $Z \sim H$.

F. An exercise. The expected life length is

$$\nu = \int_0^\infty (1 - F(t;\mu))\,dt = \int_0^\infty e^{-\mu t}\,dt = \frac{1}{\mu} \,. \tag{8.58}$$

The estimator $\hat{\nu}$ is now just the observed average life length, the empirical counterpart of ν. It is unbiased, of course. In fact, $\breve{\mu}$ and $\hat{\nu}$ are UMVUE (uniformly minimum variance unbiased estimators) since they are based on W, which is a sufficient and complete statistic. and

$$\mathbb{E}[\hat{\nu}] = \nu \;,\; n \geq 1 \,,$$

$$\mathbb{V}[\hat{\nu}] = \frac{\nu^2}{n} \;,\; n \geq 1 \,.$$

Likewise, it also holds that $\hat{\nu}$ is weakly consistent, and

$$\hat{\nu} \sim_{\text{as}} \mathrm{N}\left(\nu, \frac{1}{\mu^2 \sum_{i=1}^{n} (1 - e^{-\mu z_i})}\right) \,. \tag{8.59}$$

Copying the arguments above (or using (D.6) in Appendix D), it can also be concluded that $\hat{\nu}$ is consistent and that

$$\hat{\nu} \sim_{\text{as}} \mathrm{N}\left(\nu, \frac{1}{n\mu^2 (1 - e^{-\mu z})}\right) \,. \tag{8.60}$$

8.5 Parametric inference in the Markov model

A. The likelihood of a time-continuous Markov process. Consider now the general set-up, whereby the development of an insurance policy is represented by a continuous time Markov process Z on a finite state space $\mathcal{Z} = \{1, \ldots, J\}$. As usual, let $I_j(t)$ and $N_{jk}(t)$ denote, respectively, the indicator of the event that the process is in state j at time $t \geq 0$, and the number of transitions from state j to state k in the time interval $(0, t]$. The transition intensities $\mu_{jk}(t)$ are assumed to exist, and to be piecewise continuous.

Suppose the policy is observed continually throughout the time period $[\underline{t}, \bar{t}\,]$, commencing in state j_0 at time $\underline{t}$. One then speaks of *left-censoring* and *right-censoring* at times $\underline{t}$ and $\bar{t}$, respectively, and the triplet $z = (\underline{t}, \bar{t}, j_0)$ will be referred to as *observational design* or *censoring scheme* of the policy.

Consider a specific realization of the observed part of the process, with q transitions taking place at (or rather about) times t_p, $p = 1, \ldots, q$:

$$Z(t) = \begin{cases} j_0 & , \underline{t} < t < t_1, \\ j_1 & , t_1 + dt_1 < t < t_2, \\ & \ldots \\ j_{q-1} & , t_{q-1} + dt_{q-1} < t < t_q, \\ j_q & , t_q + dt_q < t < \bar{t}. \end{cases}$$

By the given censoring, the probability of this realization is as follows, where $t_0 = \underline{t}\,$, $t_q = \bar{t}$, and $\mu_{j\,\cdot}(t) = \sum_{k;k\neq j} \mu_{jk}(t)$ denotes the total intensity of transition out of state j at time t:

$$\exp\left(-\int_{t_0}^{t_1} \mu_{j_0\,\cdot}(t)\,dt\right)\mu_{j_0 j_1}(t_1)\,dt_1 \exp\left(-\int_{t_1}^{t_2} \mu_{j_1\,\cdot}(t)\,dt\right)\,\mu_{j_1 j_2}(t_2)\,dt_2 \cdots$$
$$\exp\left(-\int_{t_{q-1}}^{t_q} \mu_{j_{q-1}\,\cdot}(t)\,dt\right)\mu_{j_{q-1}j_q}(t_q)\,dt_q \exp\left(-\int_{t_q}^{t_{q+1}} \mu_{j_q\,\cdot}(t)\,dt\right)$$

$$\begin{aligned} &= \prod_{p=1}^{q} \mu_{j_{p-1}j_p}(t_p)dt_p \;\exp\left(-\sum_{p=1}^{q}\int_{t_{p-1}}^{t_p} \mu_{j_{p-1}\,\cdot}(t)\,dt\right) \\ &= \exp\left(\sum_{p=1}^{q} \ln \mu_{j_{p-1}j_p}(t_p)\; - \sum_{p=1}^{q}\int_{t_{p-1}}^{t_p} \mu_{j_{p-1}\,\cdot}(t)\,dt\right) dt_1 \ldots dt_{q-1}\,. \end{aligned}$$

It follows that the likelihood of the observed path of the process is

$$\begin{aligned} L &= \exp\left(\sum_{j\neq k}\int_{\underline{t}}^{\bar{t}} \ln \mu_{jk}(t)\,dN_{jk}(t) - \sum_{j}\int_{\underline{t}}^{\bar{t}} \mu_{j\,\cdot}(t) I_j(t)\,dt\right) \\ &= \exp\left(\sum_{j\neq k}\int_{\underline{t}}^{\bar{t}} \{\ln \mu_{jk}(t)\,dN_{jk}(t) - \mu_{jk}(t) I_j(t)\,dt\}\right)\,. \end{aligned} \tag{8.61}$$

B. ML estimation of parametric intensities. Now consider a parametric model where the intensities are of the form $\mu_{jk}(t, \boldsymbol{\theta})$, with $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_r)'$ varying in an open set in the r-dimensional euclidean space, $r < \infty$. We assume they are twice continuously differentiable functions of $\boldsymbol{\theta}$.

Suppose that inference is to be made about the intensities or, equivalently, the parameter θ on the basis of data from a sample of n similar policies. Equip all quantities related to the i-th policy by top-script (i). The processes Z^i are assumed to be stochastically independent replicates of the process Z described above, but their censoring schemes $z^{(i)}$ may be different.

By independence, the likelihood of the whole data set is the product of the individual likelihoods: $L = \prod_{i=1}^{n} L^{(i)}$. Thus, by (8.61),

$$\ln L = \sum_{j \neq k} \int \left(\ln \mu_{jk}(t, \boldsymbol{\theta}) \sum_{i=1}^{n} dN_{jk}^{(i)}(t) - \mu_{jk}(t, \boldsymbol{\theta}) \sum_{i=1}^{n} I_j^{(i)}(t)\, dt \right) . \tag{8.62}$$

The censoring schemes are not visualized in (8.62), and they need not be if, as a matter of definition, $dN_{jk}^{(i)}(t)$ and $I_j^{(i)}(t)$ are taken as 0 for $t \notin [\underline{t}^{(i)}, \overline{t}^{(i)}]$. Likewise, introduce

$$p_j^{(i)}(t) = p_{j_0 j}^{(i)}(\underline{t}^{(i)}, t) 1_{\,[\underline{t}^{(i)}, \overline{t}^{(i)}]}(t)\,,$$

the probability that the censored process $Z^{(i)}$ is in state j at time t, by definition taken as 0 for $t \notin [\underline{t}^{(i)}, \overline{t}^{(i)}\,]$.

In the MLE construction we need the r-vector of derivatives of (8.62),

$$\frac{\partial}{\partial \boldsymbol{\theta}} \ln L = \sum_{j \neq k} \int \frac{\partial}{\partial \boldsymbol{\theta}} \ln \mu_{jk}(t, \boldsymbol{\theta}) \left(\sum_{i=1}^{n} dN_{jk}^{(i)}(t) - \mu_{jk}(t, \boldsymbol{\theta}) \sum_{i=1}^{n} I_j^{(i)}(t) dt \right) , \tag{8.63}$$

and to find the information matrix we need the $r \times r$ matrix of second order derivatives,

$$\begin{aligned} \frac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \ln L \quad &= \quad \sum_{j \neq k} \int \frac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \ln \mu_{jk}(t, \boldsymbol{\theta}) \left(\sum_{i=1}^{n} dN_{jk}^{(i)}(t) - \mu_{jk}(t, \boldsymbol{\theta}) \sum_{i=1}^{n} I_j^{(i)}(t) dt \right) \\ & \quad - \sum_{j \neq k} \int \frac{\partial}{\partial \boldsymbol{\theta}} \ln \mu_{jk}(t, \boldsymbol{\theta}) \frac{\partial}{\partial \boldsymbol{\theta}'} \mu_{jk}(t, \boldsymbol{\theta}) \sum_{i=1}^{n} I_j^{(i)}(t)\, dt\,. \end{aligned} \tag{8.64}$$

By (8.63) the MLE $\hat{\boldsymbol{\theta}}$ is the solution to

$$\sum_{j \neq k} \int \frac{\partial}{\partial \boldsymbol{\theta}} \ln \mu_{jk}(t, \boldsymbol{\theta}) \left(\sum_{i=1}^{n} dN_{jk}^{(i)}(t) - \mu_{jk}(t, \boldsymbol{\theta}) \sum_{i=1}^{n} I_j^{(i)}(t) dt \right) |_{\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}} = \mathbf{0}^{rx1} \tag{8.65}$$

For each type of transition $j \to k$ introduce N_{jk}, the number of transitions of that type, and (if $N_{jk} > 0$) $T_{jk;1}, \ldots, T_{jk;N_{jk}}$, the times when such transitions occur. In terms of these quantities the ML equations become

$$\sum_{j \neq k} \sum_{h=1}^{N_{jk}} \frac{\frac{\partial}{\partial \theta_q} \mu_{jk}(T_{jk;h}, \hat{\boldsymbol{\theta}})}{\mu_{jk}(T_{jk;h}, \hat{\boldsymbol{\theta}})} = \sum_{j \neq k} \int \frac{\partial}{\partial \theta_q} \mu_{jk}(t, \hat{\boldsymbol{\theta}}) \sum_{i=1}^{n} I_j^{(i)}(t) dt\,, \tag{8.66}$$

$q = 1, \ldots, r$. The form (8.66) is explicit and is the one we will work with when it comes to numerical computation of the MLE: The good thing about the form (8.65) is that it expresses the log likelihood as the sum of contributions from all small time intervals. This is particularly useful in the derivation of the information matrix as was seen already in Paragraph 8.3.A.

Referring to Appendix D, the large sample distribution properties of the MLE are given by

$$\hat{\boldsymbol{\theta}} \sim_{\text{as}} \mathrm{N}(\boldsymbol{\theta}, \boldsymbol{\mathcal{J}}_n(\boldsymbol{\theta})^{-1}), \tag{8.67}$$

where $\boldsymbol{\mathcal{J}}_n(\boldsymbol{\theta})$ is the information matrix,

$$\boldsymbol{\mathcal{J}}_n(\boldsymbol{\theta}) = -\mathbb{E}\left[\frac{\partial}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'} \ln L\right]. \tag{8.68}$$

Taking expectation in (8.64), noting that the terms

$$dN_{jk}^{(i)}(t) - \mu_{jk}(t,\boldsymbol{\theta})I_j^{(i)}(t)dt$$

all have zero mean, we obtain

$$\boldsymbol{\mathcal{J}}_n(\boldsymbol{\theta}) = \sum_{j\neq k} \int \frac{1}{\mu_{jk}(t,\boldsymbol{\theta})} \left(\frac{\partial}{\partial\boldsymbol{\theta}}\mu_{jk}(t,\boldsymbol{\theta}) \frac{\partial}{\partial\boldsymbol{\theta}'}\mu_{jk}(t,\boldsymbol{\theta})\right) \sum_{i=1}^{n} p_j^{(i)}(t,\boldsymbol{\theta})dt\,. \tag{8.69}$$

The expression in parentheses under the integral sign is an $r \times r$ matrix and all other quantities are scalar.

It is seen that the information matrix tends to infinity, hence the variance matrix of the MLE tends to 0, if the terms $\sum_{i=1}^{n} p_j^{(i)}(t,\boldsymbol{\theta})$ grow to infinity as n increases, roughly speaking, which means that the expected number of individuals exposed to risk in different states gets unlimited.

8.6 Piece-wise constant intensities

A. Piece-wise constant intensities. Let $\underline{t} = t_0 < t_1 < \cdots < \cdots < t_m = \bar{t}$ be some finite partition of the time interval $[\underline{t}, \bar{t}\,]$, and assume that the intensities are constant in each interval $[t_{q-1}, t_q)$, that is, they are step functions of the form

$$\mu_{jk}(t) = \sum_{q=1}^{m} 1_{[t_{q-1},t_q)}(t)\mu_{jk,q}\,, \tag{8.70}$$

where the $\mu_{jk,q}$ take values in $(0,\infty)$, with no relationships between them. The situation fits into the general framework with $\boldsymbol{\theta} = (\ldots, \mu_{jk,q}, \ldots)'$, a vector of (typically high) dimension.

B. The MLE estimators are O-E rates. The log likelihood in (8.62) now becomes

$$\ln L = \sum_{j\neq k}\sum_{q=1}^{m}\{\ln \mu_{jk,q} N_{jk,q} - \mu_{jk,q} W_{j,q}\}\,, \tag{8.71}$$

where

$$N_{jk,q} = \int_{t_{q-1}}^{t_q} \sum_{i=1}^{n} dN_{jk}^{(i)}(t)\,, \tag{8.72}$$

$$W_{j,q} = \int_{t_{q-1}}^{t_q} \sum_{i=1}^{n} I_j^{(i)}(t)\,dt\,, \tag{8.73}$$

are, respectively, the total number of transitions from state j to state k and the total time spent in state j during the age interval $[t_{q-1}, t_q)$.

Since the $\mu_{jk,q}$ are functionally unrelated, the log likelihood decomposes into terms that depend on one and only one of the basic parameters, and finding maximum amounts to maximizing each term. The derivatives involved in the ML construction now become particularly simple:

$$\frac{\partial}{\partial \mu_{jk,q}} \ln L = \frac{1}{\mu_{jk,q}} N_{jk,q} - W_{j,q}\,, \tag{8.74}$$

$$\frac{\partial^2}{\partial \mu_{jk,q} \partial \mu_{j'k',q'}} \ln L = -\delta_{jkq,j'k'q'} \frac{1}{\mu_{jk,q}^2} N_{jk,q}\,. \tag{8.75}$$

It follows from (8.74) that the MLE is

$$\hat{\mu}_{jk,q} = \frac{N_{jk,q}}{W_{j,q}}\,, \tag{8.76}$$

an O-E rate of the same kind as in the simple model of Section 8.4. Noting that, by (8.72),

$$\mathbb{E}[N_{jk,q}] = \mu_{jk,q} \int_{t_{q-1}}^{t_q} \sum_{i=1}^{n} p_j^{(i)}(t)\,dt\,,$$

we obtain from (8.75) that the information matrix becomes

$$\boldsymbol{\mathcal{J}}_n(\boldsymbol{\theta}) = \mathrm{Diag}\left(\ldots, \frac{\int_{t_{q-1}}^{t_q} \sum_{i=1}^{n} p_j^{(i)}(t)\,dt}{\mu_{jk,q}}, \ldots \right), \tag{8.77}$$

a diagonal matrix, implying that the estimators of the $\mu_{jk,q}$ are asymptotically independent.

An estimator of the information is obtained upon replacing the parameter functions appearing on the right of (8.77) by their straightforward estimators: put $\mu_{jk,q} \approx \hat{\mu}_{jk,q}$ defined by (8.76) and

$$\int_{t_{q-1}}^{t_q} \sum_{i=1}^{n} p_j^{(i)}(t)\,dt = \mathbb{E}[W_{j,q}] \approx W_{j,q}\,,$$

to obtain the estimate

$$\hat{\boldsymbol{\mathcal{J}}}_n = \mathrm{Diag}\left(\ldots, \frac{W_{j,q}^2}{N_{jk,q}}, \ldots \right). \tag{8.78}$$

C. Smoothing O-E rates. The MLE of the intensity function is obtained upon inserting the estimators (8.76) in (8.70). The resulting function will typically have a ragged appearance due to the estimation error in a finite sample. This is unsatisfactory since the intensities are expected to be smooth functions: for instance, there are a priori reasons to assume that the mortality intensity is a continuous and non-decreasing function of the age. Now, the very assumption of piece-wise constant intensities is artificial, of course, and the estimates obtained under this assumption cannot serve as an ultimate answer in practice. In fact, they represent only the first step in a two-stage procedure, where the second step is to fit some smooth functions to the raw estimates delivered by the O-E rates. The functions used for fitting constitute the model we have in mind. It may be objected that the two-stage procedure is a detour since, if the intensities are assumed to be functions of a smaller set of parameters, one could follow the prescription in Section 8.2 and maximize the likelihood directly. There are two reasons why the two-stage procedure never the less merits special treatment: in the first place, the O-E rates and their asymptotic variance matrix are easy to construct; in the second place, a comparative plot of the fitted functions and the O-E rates makes it possible to detect systematic deviations between model assumptions and facts.

A commonly used fitting technique is the so-called generalized least squares method, which amounts to minimizing a positive definite quadratic form in the deviations between the raw estimates and the fitting functions. Let us focus on one intensity that is to be graduated and, to fix ideas, assume it is the mortality intensity in the simple model with two states 'alive' and 'dead'.

The ML estimators are the occurrence-exposure rates,

$$\hat{\mu}_q = \frac{N_q}{W_q}\,,$$

which are well defined for all q such that $W_q > 0$ (i.e. in age intervals where there were survivors exposed to risk of death). The $\hat{\mu}_q$ are asymptotically (as n increases) normally distributed, mutually independent, unbiased, and with variances given by

$$\sigma_q^2 = \text{as.}\mathbb{V}[\hat{\mu}_q] = \frac{\mu_q}{\mathbb{E}[W_q]}\,, \tag{8.79}$$

where the expected exposure is

$$\mathbb{E}[W_q] = \sum_{i=1}^{n} \int_{q-1}^{q} p^{(i)}(t)\,dt\,,$$

$p^{(i)}(t)$ being the probability that individual No. i is alive and under observation at time t.

The variance σ_q^2 is inversely proportional to the corresponding expected exposure. In the present case, with only one intensity of transition from the state 'alive' to the absorbing state 'dead', we find explicit expressions for the expected exposure.

For instance, suppose we have observed each individual life from birth until death or until attained age 100, whichever occurs first (i.e. censoring at age 100). Then, for $t \in [q-1, q)$ with $q = 1, \ldots, 100$, we have

$$\begin{aligned} p^{(i)}(t) &= \exp\left(-\int_0^t \mu(s)\,ds\right) \\ &= \exp\left(-\sum_{p=1}^{q-1} \mu_p - (t-(q-1))\,\mu_q\right), \end{aligned} \tag{8.80}$$

hence

$$\begin{aligned} \mathbb{E}[W_q] &= n \int_{q-1}^{q} \exp\left(-\sum_{p=1}^{q-1} \mu_p - (t-(q-1))\,\mu_q\right) dt \\ &= n \exp\left(-\sum_{p=1}^{q-1} \mu_p\right) \frac{1-\exp\left(-\mu_q\right)}{\mu_q}, \end{aligned}$$

and

$$\sigma_q^2 = \frac{1}{n} \frac{\mu_q}{\exp\left(-\sum_{p=1}^{q-1} \mu_p\right)\left(1-\exp\left(-\mu_q\right)\right)}. \tag{8.81}$$

You should look at other censoring schemes and discuss the impact of censoring on the variance. Take e.g. the case where person No i enters at age x_i and is observed until death or age 100, whichever occurs first (all x_i less than 100).

Estimators $\hat{\sigma}_q^2$ of the variances are obtained upon replacing the μ_j in (8.81) by the estimators $\hat{\mu}_j$. Simpler estimators are obtained by just replacing μ_q and $\mathbb{E}[W_q]$ in (8.79) with their straightforward empirical counterparts: $\hat{\sigma}_q^2 = \hat{\mu}_q / W_q = N_q / W_q^2$.

Now to graduation. The occurrence-exposure rates will usually have a ragged appearance. Assuming that the real underlying mortality intensity is a smooth function, we will therefore fit a suitable function to the occurrence-exposure rates. Suppose we assume that the true mortality rate is a Gompertz-Makeham function, $\mu(t) = \alpha + \beta e^{\gamma t}$. Then, take some representative age t_q (typically $t_q = q - 0.5$) in each age interval and fit the parameters α, β, γ by minimizing a weighted sum of squared errors

$$Q = \sum_q a_q \left(\hat{\mu}_q - \alpha - \beta e^{\gamma t_q}\right)^2.$$

This is a matter of non-linear regression. The optimal weights a_q are the inverse of the variances, but since these are unknown, we plug in the estimators and use $a_q = 1/\hat{\sigma}_q^2$.

The minimizing values α^*, β^*, and γ^* are obtained by differentiating Q with respect to each of the three parameters and setting the derivatives equal to 0.

The derivatives are:

$$\begin{aligned}
\frac{\partial}{\partial\alpha}Q &= \sum_q a_q 2\left(\hat{\mu}_q - \alpha - \beta e^{\gamma t_q}\right)(-1)\,,\\
\frac{\partial}{\partial\beta}Q &= \sum_q a_q 2\left(\hat{\mu}_q - \alpha - \beta e^{\gamma t_q}\right)(-e^{\gamma t_q})\,,\\
\frac{\partial}{\partial\gamma}Q &= \sum_q a_q 2\left(\hat{\mu}_q - \alpha - \beta e^{\gamma t_q}\right)(-\beta e^{\gamma t_q} t_q)\,.
\end{aligned}$$

Thus α^*, β^*, and γ^* are the solution to the equations

$$\begin{aligned}
\sum_q a_q \left(\hat{\mu}_q - \alpha^* - \beta^* e^{\gamma^* t_q}\right) &= 0\,,\\
\sum_q a_q \left(\hat{\mu}_q - \alpha^* - \beta^* e^{\gamma^* t_q}\right) e^{\gamma^* t_q} &= 0\,,\\
\sum_q a_q \left(\hat{\mu}_q - \alpha^* - \beta^* e^{\gamma^* t_q}\right) e^{\gamma^* t_q} t_q &= 0\,.
\end{aligned}$$

This is in general a set of non-linear equations that does not allow of an explicit solution. Actually these equations are just as involved as the maximum likelihood equations in Paragraph 9.2C above, which is disappointing since the two-stage procedure considered here was supposed to be simpler. (Occurrence-exposure rates are easy to find and they are asymptotically independent, which makes it easy to find their asymptotic variances. The graduation is, however, messy.)

Suppose now that γ is taken to be known. Then only the two first equations above are relevant and they reduce to

$$\begin{aligned}
\sum_q a_q\,\alpha^* + \sum_q a_q e^{\gamma t_q}\,\beta^* &= \sum_q a_q \hat{\mu}_q\,,\\
\sum_q a_q e^{\gamma^* t_q}\,\alpha^* - \sum_q a_q e^{2\gamma^* t_q}\,\beta^* &= \sum_q a_q \hat{\mu}_q e^{\gamma^* t_q}\,.
\end{aligned}$$

This is a linear system of equations with an explicit solution, which the industrious reader certainly will find.

For each interval $[t_{q-1}, t_q)$ choose a "representative" point t_q, e.g. the interval midpoint. Put $\hat{\mu} = (\ldots, \hat{\mu}_q, \ldots)'$, the vector of O-E rates, and (with a bit sloppy notation) $\mu(\boldsymbol{\theta}) = (\ldots, \mu(t_q, \boldsymbol{\theta}), \ldots)'$, the vector of true values. Let $A = (a_{pq})$ be some positive definite matrix of order $r \times r$. Estimate $\boldsymbol{\theta}$ by $\boldsymbol{\theta}^*$ minimizing

$$(\mu(\boldsymbol{\theta}) - \hat{\mu})' A (\mu(\boldsymbol{\theta}) - \hat{\mu}) = \sum_{pq} a_{pq}(\mu(t_p, \boldsymbol{\theta}) - \hat{\mu}_p)(\mu(t_q, \boldsymbol{\theta}) - \hat{\mu}_q)\,. \tag{8.82}$$

If the intensity is a linear function of $\boldsymbol{\theta}$ (like in the G-M study with known γ),

$$\mu(\boldsymbol{\theta}) = Y(t)\boldsymbol{\theta}\,, \tag{8.83}$$

then

$$\hat{\boldsymbol{\theta}} = (Y'AY)^{-1}Y'A\hat{\mu}\,. \tag{8.84}$$

The asymptotic variance of $\hat{\boldsymbol{\theta}}$ is $(Y'AY)^{-1}Y'A\boldsymbol{\Sigma}(\boldsymbol{\theta})AY(Y'AY)^{-1}$. By the Gauss-Markov theorem it is minimized by taking $A = \boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1}$, and the minimum is $(Y'\boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1}Y)^{-1}$. Thus, asymptotically the best choice of A is $\hat{\boldsymbol{\Sigma}}^{-1}$, where $\hat{\boldsymbol{\Sigma}}$ is some estimate of $\boldsymbol{\Sigma}$ satisfying (8.87). Write $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}(\boldsymbol{\theta})$. The symmetric, pd matrix $\boldsymbol{\Sigma}$ has a symmetric pd square root W such that $\boldsymbol{\Sigma} = W^2$.

$$\begin{aligned}
\Delta &= (Y'AY)^{-1}Y'A\boldsymbol{\Sigma}AY(Y'AY)^{-1} - (Y'\boldsymbol{\Sigma}^{-1}Y)^{-1} \\
&= (Y'AY)^{-1}Y'AW\left[I - WY((WY)'(WY))^{-1}(WY)'\right]WAY(Y'AY)^{-1} \\
&= (Y'AY)^{-1}Y'AW\,H\,WAY(Y'AY)^{-1}\,.
\end{aligned}$$

where

$$H = I - P(P'P)^{-1}P'\,.$$

The matrix $H = I - P(P'P)^{-1}P'$ is symmetric, $H = H'$, and idempotent, $H^2 = H$, Thus, $H = H'H$ and

$$\Delta = \left(HWAY(Y'AY)^{-1}\right)'\left(HWAY(Y'AY)^{-1}\right)$$

which is indeed pd.

8.7 Impact of the censoring scheme

A. The precision of the estimation. The precision of the MLE depends on the amount of information provided by the censoring scheme of the study. Asymptotically it is the variance matrix $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ that determines everything, and in Section 8.2 it was pointed out that the size of this matrix depends on the censoring scheme only through the functions $\sum_{i=1}^{n} p_j^{(i)}(t)$, $j = 0, \ldots, J$, the expected numbers of individuals staying in each state g at time t. (It depends also on the parametric structure of the intensities, of course.) We shall look at two censoring schemes frequently encountered in practice.

B. Longitudinal observation (cohort studies). The term cohort stems from Latin and originally signified a unit division in an ancient Roman legion. In demography it means a class of individuals born in a particular year or more general period of time (a "generation"), and a cohort study is one where a cohort is observed over a certain period, possibly until it is extinct. This was the situation in Paragraph 8.4.B.

Thus, let the n Markov processes in the general set-up be stochastic replicates, all commencing in state 0 at time 0 and thereafter observed continuously throughout the time interval $[0, \bar{t}\,]$. In this case

$$\sum_{i=1}^{n} p_j^{(i)}(t) = n\,p_1^{(i)}(t)\,,\; j = 0, \ldots, J,$$

and

$$\boldsymbol{\mathcal{J}}_n(\boldsymbol{\theta}) = n \sum_{j \neq k} \int_0^{\bar{t}} \frac{1}{\mu(t,\boldsymbol{\theta})} \frac{\partial}{\partial \boldsymbol{\theta}} \mu(t,\boldsymbol{\theta}) \frac{\partial}{\partial \boldsymbol{\theta}'} \mu(t,\boldsymbol{\theta}) p_1^{(i)}(t) dt \,.$$

The inverse of this matrix tends to 0 as n increases.

C. Cross-sectional observation. In a cross-sectional study a population is observed over a certain period of time. As an example, suppose the G-M mortality study in Paragraph 8.2.C is conducted cross-sectionally throughout a calendar period of duration $\bar{t}$, and that it comprises n individuals at ages $\underline{t}^{(i)}$, $i = 1, \ldots, n$, at the beginning of the study. In this case the factor depending on the design in the information matrix is

$$\sum_{i=1}^{n} p^{(i)}(t) = \sum_{i=1}^{n} 1_{[\,\underline{t}^{(i)}, \underline{t}^{(i)} + \bar{t}\,]}(t) \exp\left(-\alpha(t - \underline{t}^{(i)}) - \beta \frac{e^{\gamma \underline{t}^{(i)}}(e^{\gamma t} - 1)}{\gamma} \right).$$

8.8 Confidence regions

A. An asymptotic confidence ellipsoid. From the asymptotic normality of the MLE it follows that

$$(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})' \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \sim_{\text{as}} \chi^2_r \,, \tag{8.85}$$

the chi-squared distribution with r degrees of freedom. Therefore, denoting the $(1-\varepsilon)$-fractile of this distribution by $\chi^2_{r,\,1-\varepsilon}$, an asymptotic $1-\varepsilon$ confidence region is the set of all $\boldsymbol{\theta}$ satisfying

$$(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})' \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}) \leq \chi^2_{r,\,1-\varepsilon} \,. \tag{8.86}$$

The expression on the left here will typically be a complicated function of $\boldsymbol{\theta}$, and it is in general not easy to find the values of $\boldsymbol{\theta}$ that satisfy the inequality and constitute a confidence region. Now, suppose $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ can be estimated by some function of the data, $\hat{\boldsymbol{\Sigma}}$, and that the estimator is consistent in the sense that

$$\hat{\boldsymbol{\Sigma}} \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}) \rightarrow I \,. \tag{8.87}$$

Then it is easy to show that also the relation

$$(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})' \hat{\boldsymbol{\Sigma}}^{-1} (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}) \leq \chi^2_{r,\,1-\varepsilon} \tag{8.88}$$

determines an asymptotic $1-\varepsilon$ confidence region. The relation (8.88) defines an ellipsoid, which is a fairly simple geometric figure and, as we shall see in the following paragraph, a convenient basis for deriving other confidence statements of interest.

A straightforward way of constructing $\hat{\mathbf{\Sigma}}$ would be to replace $\boldsymbol{\theta}$ in $\mathbf{\Sigma}(\boldsymbol{\theta})$ by the consistent estimator $\hat{\boldsymbol{\theta}}$, that is, put

$$\hat{\mathbf{\Sigma}} = \mathbf{\Sigma}(\hat{\boldsymbol{\theta}})\,.$$

This works well if the entries in $\mathbf{\Sigma}(\boldsymbol{\theta})$ are closed expressions in $\boldsymbol{\theta}$. Unfortunately, this is the case only in certain simple situations, typically when the state space $\mathcal{Z}$ is small and the pattern of transitions is hierarchical. One example is the mortality study with parametric mortality law, e.g. of G-M type. In more complex situations we cannot in general find closed formulas for the probabilities $p_g^{(i)}(t)$ involved in $\mathbf{\Sigma}$, even if the intensities themselves are simple parametric functions. Then a different construction is required. A simple device is to replace the $p_g^{(i)}(t)$ by their empirical counterparts $I_j^{(i)}(t)$ and put

$$\sum_{i=1}^{n} p_g^{(i)}(t) \approx I_j(t)\,. \tag{8.89}$$

B. Simultaneous confidence intervals. The confidence ellipsoid (8.88) can be resolved in simultaneous confidence intervals for all linear functions of $\boldsymbol{\theta}$ in the following way. The Schwarz inequality says that for all vectors a and x in $\mathcal{R}^s$,

$$|a'x| \le \sqrt{a'a\,x'x}\,,$$

with equality for $a = cx$. Thus, noting that the quadratic form on the left of (8.88) is $(\hat{\mathbf{\Sigma}}^{-1/2}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}))'\,(\hat{\mathbf{\Sigma}}^{-1/2}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}))$, the confidence statement can be cast equivalently as

$$|a'\hat{\mathbf{\Sigma}}^{-1/2}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})| \le \sqrt{a'a\,\chi^2_{r,\,1-\varepsilon}}\,,\ \forall a\,. \tag{8.90}$$

Since $\hat{\mathbf{\Sigma}}$ is of full rank, the vector $\hat{\mathbf{\Sigma}}^{-1/2}a$ ranges through all of $\mathcal{R}^s$ as a ranges in $\mathcal{R}^s$. Thus, writing $a'a = (\hat{\mathbf{\Sigma}}^{-1/2}a)'\hat{\mathbf{\Sigma}}(\hat{\mathbf{\Sigma}}^{-1/2}a)$, (8.90) is equivalent to

$$|a'(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})| \le \sqrt{a'\hat{\mathbf{\Sigma}}a\,\chi^2_{r,\,1-\varepsilon}}\,,\ \forall a\,,$$

that is,

$$a'\boldsymbol{\theta} \in [a'\hat{\boldsymbol{\theta}} - \sqrt{\chi^2_{r,\,1-\varepsilon}a'\hat{\mathbf{\Sigma}}a}\,,\ a'\hat{\boldsymbol{\theta}} + \sqrt{\chi^2_{r,\,1-\varepsilon}a'\hat{\mathbf{\Sigma}}a}]\,,\ \forall a\,. \tag{8.91}$$

The intervals in (8.91) are (asymptotic) simultaneous confidence intervals for all linear functions of $\boldsymbol{\theta}$ in the sense that the probability is at least $1 - \varepsilon$ that they all hold true.

C. Confidence band for the G-M mortality intensity. Returning to the mortality study example in Paragraph 8.2.C, let γ be taken as known so that the mortality intensity is a linear function of the unknown parameter $\boldsymbol{\theta} = (\alpha, \beta)'$. The MLE is obtained by solving the equations (8.35) and (8.35), and the appropriate variance matrix $\mathbf{\Sigma}$ is obtained by inverting the upper left 2×2 block in the information matrix defined by (8.69), (8.35), and (8.36).

From (8.91) we obtain simultaneous confidence intervals for all $\mu(t) = \alpha + \beta e^{\gamma t}$, constituting a confidence band in the space of mortality intensity functions;

$$\mu(t) \in [\hat{\alpha} + \hat{\beta} e^{\gamma t} - \sqrt{\chi^2_{2,\,1-\varepsilon} \hat{\sigma}_t}\,,\, \hat{\alpha} + \hat{\beta} e^{\gamma t} - \sqrt{\chi^2_{2,\,1-\varepsilon} \hat{\sigma}_t}]\,,\ \forall t > 0\,, \tag{8.92}$$

where

$$\hat{\sigma}_t = (1,\, e^{\gamma t}) \hat{\mathbf{\Sigma}} \begin{pmatrix} 1 \\ e^{\gamma t} \end{pmatrix} .$$

8.9 More on simultaneous confidence intervals

B. Confidence ellipsoid for a normal mean; Scheffé intervals.
Let $\hat{\boldsymbol{\theta}}$ is an estimator of an r-dimensional parameter vector $\boldsymbol{\theta}$ and assume that

$$\hat{\boldsymbol{\theta}} \sim N(\boldsymbol{\theta}, \mathbf{\Sigma})\,,$$

with $\mathbf{\Sigma}$ known. A $1 - \varepsilon$ confidence region of $\boldsymbol{\theta}$ is the ellipsoid defined by (3.2):

$$\mathcal{C} = \{\boldsymbol{\theta}\,;\, (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})' \mathbf{\Sigma}^{-1} (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}) \leq \chi^2_{r,\,1-\varepsilon}\}\,.$$

To construct the confidence interval (8.2) for a specific function $g(\boldsymbol{\theta})$, we are left with the mathematical problem of finding the extrema of g over the ellipsoid, which may be a difficult task. For linear functions it is simple, however, and it goes by the the technique in Paragraph 3B, which is due to Scheffé. The simultaneous confidence intervals for linear functions $a'\boldsymbol{\theta} = \sum_{p=1}^{s} a_p \boldsymbol{\theta}_p$ are, with a bit sloppy notation,

$$a'\boldsymbol{\theta} \in a'\hat{\boldsymbol{\theta}} \pm \sigma_a \sqrt{\chi^2_{r,\,1-\varepsilon}}\,,\ \forall a \in \mathcal{R}^s\,,$$

where

$$\sigma_a^2 = a'\,\mathbf{\Sigma}\,a$$

is the variance of the point estimator $a'\hat{\boldsymbol{\theta}}$.

C. Narrowing the confidence intervals.
Generally speaking, in the presence of uncertainty, the price we have to pay for making many safe statements is that each individual statement has to be vague. In our situation this general truth takes a very manifest form: for a fixed confidence level $1 - \varepsilon$ the lengths of the confidence intervals increase with the dimension of $\boldsymbol{\theta}$ since $\chi^2_{r,\,1-\varepsilon}$ is an increasing function of r (why?).

We can gain precision in terms of lengths of the intervals by reducing the number of statements we want to make. Suppose we are only interested in drawing inferences about linear combinations of r linearly independent functions

$b_j'\boldsymbol{\theta}$, $j = 1, \ldots, r$, $r < s$. Thus, putting $B = (b_1, \ldots, b_r)$, an $s \times r$ matrix, we are only interested in linear functions $a'\boldsymbol{\theta}$ with $a = Bc$ for some r-vector c, that is, $a \in \mathcal{R}(B)$, the r-dimensional linear space spanned by the b_j. Then, start from

$$B'\hat{\boldsymbol{\theta}} \sim N(B'\boldsymbol{\theta}, B'\boldsymbol{\Sigma}B)\,,$$

and apply the results above to obtain that simultaneous confidence intervals for all linear functions of $B'\,\boldsymbol{\theta}$ are given by

$$c'B'\boldsymbol{\theta} \in c'B'\hat{\boldsymbol{\theta}} \pm \sqrt{c'B'\boldsymbol{\Sigma}Bc\,\chi^2_{r,\,1-\varepsilon}}\,,\ \forall\, c \in \mathcal{R}^r\,, \tag{8.93}$$

or, equivalently,

$$a'\boldsymbol{\theta} \in a'\hat{\boldsymbol{\theta}} \pm \sigma_a\sqrt{\chi^2_{r,\,1-\varepsilon}}\,,\ \forall a \in \mathcal{R}(B)\,.$$

It is seen that, by reducing the "dimension of our statements" from s to r, we have gained a reduction of the lengths of the confidence intervals by a factor $\sqrt{\chi^2_{r,\,1-\varepsilon}/\chi^2_{r,\,1-\varepsilon}}$.

B. Bonferroni intervals for a finite number of parameter functions. Let $g_\ell(\boldsymbol{\theta})$, $\ell = 1, \ldots, m$, be a finite number of real-valued parameter functions, and assume that for each $g_\ell(\boldsymbol{\theta})$ we have constructed an individual confidence interval $[\,\underline{g}_j\,,\,\overline{g}_j\,]$ with level $1-\varepsilon_j$. Thus, denoting the event $g_\ell(\boldsymbol{\theta}) \in [\underline{g}_j\,,\,\overline{g}_j]$ by A_j, we have $P_{\boldsymbol{\theta}}(A_j) \geq 1-\varepsilon_j$ for each j. The probability that all A_j hold true at the same time is

$$\begin{aligned}
P_{\boldsymbol{\theta}}\left(\cap_j A_j\right) &= 1 - P_{\boldsymbol{\theta}}\left(\cup_j A_j^c\right) \\
&\geq 1 - \sum_{j=1}^{r} P_{\boldsymbol{\theta}}(A_j^c) \\
&\geq 1 - \sum_{j=1}^{r} \varepsilon_j\,.
\end{aligned}$$

It follows that the intervals taken together are simultaneous confidence intervals with simultaneous confidence level no less than $1-\varepsilon$, where $\varepsilon = \sum_{j=1}^{r}\varepsilon_j$.

This simple device, due to Bonferroni, represents an attractive alternative to the approach in Paragraphs A – C in situations where we take interest only in a finite number of parameter functions. It turns out that the Bonferroni intervals often are shorter than the Scheffé intervals, which aim at an infinite number of parameter functions. Bonferroni intervals with simultaneous confidence level $1-\varepsilon$ for q linear functions $a_j'\boldsymbol{\theta}$, $j = 1, \ldots, q$, are

$$a_j'\boldsymbol{\theta} \in a_j'\hat{\boldsymbol{\theta}} \pm \sigma_{a_j}\sqrt{\chi^2_{1,\,1-\varepsilon/q}}\,, \tag{8.94}$$

(Note that $\sqrt{\chi^2_{1,\,1-\varepsilon/q}}$ is just the $(1-\varepsilon/2q)$-fractile of the standard normal distribution, so we recognize (8.94) as the traditional individual $1-\varepsilon/q$ confidence interval for a normal mean.) Let r ($\leq s$) be the dimension of the space spanned by the coefficient vectors a_j. The corresponding Scheffé intervals based on (8.93) are

$$a_j'\boldsymbol{\theta} \in a_j'\hat{\boldsymbol{\theta}} \pm \sigma_{a_j}\sqrt{\chi^2_{r,\,1-\varepsilon}}\,.$$

The ratio of the length of the intervals by the two constructions is $B/S(q,r,\varepsilon) = \sqrt{\chi^2_{1,\,1-\varepsilon/q}/\chi^2_{r,\,1-\varepsilon}}$. Clearly, the ratio decreases with r. It increases with q, and (for $r > 1$) it starts from $q = 1$ with a value smaller than 1 and tends to infinity as q grows. There will be some value $q(r,\varepsilon)$ such that the ratio is ≤ 1 for $q \leq q(r,\varepsilon)$. Inspection of a table of the chi-square fractiles shows e.g. that $q(2, 0.025) = 5$, $q(4, 0.1) = 20$, $q(6, 0.25) = 50$ (approximately).

E. Confidence intervals based on consistent and asymptotically normal point estimators.

The results and considerations in the previous paragraphs carry over to the situation in Section 3, where (3.4) formed the basis for simultaneous inference.

Suppose we are interested in functions of a set of parameter functions $f_j(\boldsymbol{\theta})$, $\ell = 1,\ldots,q$, $q < r$. Put $f = (f_1,\ldots,f_q)'$. If f is continuously differentiable, first order Taylor expansion gives

$$f(\hat{\boldsymbol{\theta}}) \sim_{\text{as}} N\left(f(\boldsymbol{\theta}),\, \boldsymbol{\Sigma}_f(\boldsymbol{\theta})\right)\,,$$

with

$$\boldsymbol{\Sigma}_f(\boldsymbol{\theta}) = Df(\boldsymbol{\theta})'\,\boldsymbol{\Sigma}(\boldsymbol{\theta})\,Df(\boldsymbol{\theta})\,,$$

and

$$Df(\boldsymbol{\theta}) = \frac{\partial}{\partial\boldsymbol{\theta}}f(\boldsymbol{\theta})\,,$$

an $r \times q$ matrix. We obtain the asymptotic confidence ellipsoid

$$\mathcal{C}_f = \{f\,;(f - f(\hat{\boldsymbol{\theta}}))'\hat{\boldsymbol{\Sigma}}_f^{-1}(f - f(\hat{\boldsymbol{\theta}})) \leq \chi^2_{q,\,1-\varepsilon}\}\,,$$

where $\hat{\boldsymbol{\Sigma}}_f$ is some consistent estimator of $\boldsymbol{\Sigma}_f$ in the sense of (3.3), e.g. $\hat{\boldsymbol{\Sigma}}_f = \boldsymbol{\Sigma}_f(\hat{\boldsymbol{\theta}})$. Asymptotic Scheffé intervals for all functions $g(\boldsymbol{\theta}) = h(f(\boldsymbol{\theta}))$, with h real-valued and continuously differentiable, are

$$g(\boldsymbol{\theta}) \in g(\hat{\boldsymbol{\theta}}) \pm \hat{\sigma}_g\sqrt{\chi^2_{q,\,1-\varepsilon}}\,,$$

where

$$\hat{\sigma}_g = \frac{\partial}{\partial\boldsymbol{\theta}'}g(\boldsymbol{\theta})\,\boldsymbol{\Sigma}(\boldsymbol{\theta})\,\frac{\partial}{\partial\boldsymbol{\theta}}g(\boldsymbol{\theta})\,|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}\,.$$

Asymptotic Bonferroni intervals for a finite collection of functions g_ℓ, $\ell = 1,\ldots,q$, are obtained upon replacing r and ε with 1 and ε/q.

F. The G-M mortality intensity revisited.
The confidence intervals (3.8) are infinitely many, so Bonferroni ideas cannot help here. If also c is to be estimated, we obtain asymptotic confidence intervals by the device above. The same goes for any function of actuarial relevance, like the reserve of a life insurance or a portfolio of insurances. Think of examples.
Returning to the mortality study example in Paragraph 2C, let γ be taken as known so that the mortality intensity is a linear function of the unknown parameter $\boldsymbol{\theta} = (\alpha, \beta)'$. The MLE is obtained by solving the eqnarrays (8.35) and (8.35), and the appropriate variance matrix $\boldsymbol{\Sigma}$ is obtained by inverting the upper left 2×2 block in the information matrix defined by (8.69), (8.35), and (8.36).

From (8.91) we obtain simultaneous confidence intervals for all $\mu(t) = \alpha + \beta e^{\gamma t}$, constituting a confidence band in the space of mortality intensity functions;

$$\mu(t) \in [\hat{\alpha} + \hat{\beta} e^{\gamma t} - \sqrt{\chi^2_{2,\,1-\varepsilon} \hat{\sigma}_t}\,,\ \hat{\alpha} + \hat{\beta} e^{\gamma t} - \sqrt{\chi^2_{2,\,1-\varepsilon} \hat{\sigma}_t}]\,,\ \forall t > 0\,,$$

where

$$\hat{\sigma}_t = (1,\, e^{\gamma t}) \hat{\boldsymbol{\Sigma}} \begin{pmatrix} 1 \\ e^{\gamma t} \end{pmatrix}.$$

Chapter 9

Safety loadings and bonus

9.1 General considerations

A. Bonus – what it is. The word *bonus* is Latin and means 'good'. In insurance terminology it denotes various forms of repayments to the policyholders of that part of the company's surplus that stems from good performance of the insurance portfolio, a sub-portfolio, or the individual policy. We shall here concentrate on the special form it takes in traditional life insurance.

The issue of bonus presents itself in connection with every *standard* life insurance contract, characteristic of which is its specification of nominal contingent payments that are binding to both parties throughout the term of the contract. All contracts discussed so far are of this type, and a concrete example is the combined policy described in 7.4: upon inception of the contract the parties agree on a death benefit of 1 and a disability benefit of 0.5 per year against a level premium of 0.013108 per year, regardless of future developments of the intensities of mortality, disability, and interest. Now, life insurance policies like this one are typically long term contracts, with time horizons wide enough to capture significant variations in intensities, expenses, and other relevant economic-demographic conditions. The uncertain development of such conditions subjects every supplier of standard insurance products to a risk that is non-diversifiable, that is, independent of the size of the portfolio; an adverse development can not be countered by raising premiums or reducing benefits, and also not by cancelling contracts (the right of withdrawal remains one-sidedly with the insured). The only way the insurer can safeguard against this kind of risk is to build into the contractual premium a safety loading that makes it cover, on the average in the portfolio, the contractual benefits under any likely economic-demographic development. Such a safety loading will typically create a systematic surplus, which by statute is the property of the insured and has to be repaid in the form of bonus.

B. Sketch of the usual technique. The approach commonly used in practice is the following. At the outset the contractual benefits are valuated, and the premium is set accordingly, on a *first order (technical) basis*, which is a set of hypothetical assumptions about interest, intensities of transition between policy-states, costs, and possibly other relevant technical elements. The first order model is a means of prudent calculation of premiums and reserves, and its elements are therefore placed to the safe side in a sense that will be made precise later. As time passes reality reveals true elements that ultimately set the realistic scenario for the entire term of the policy and constitute what is called the *second order (experience) basis*. Upon comparing elements of first and second order, one can identify the safety loadings built into those of first order and design schemes for repayment of the systematic surplus they have created. We will now make these things precise.

To save notation, we disregard administration expenses for the time being and discuss them separately in Section 9.7 below.

9.2 First and second order bases

A. The second order model. The policy-state process Z is assumed to be a time-continuous Markov chain as described in Section 7.2. In the present context we need to equip the indicator processes and counting processes related to the process Z with a topscript, calling them I_j^Z and N_{jk}^Z. The probability measure and expectation operator induced by the transition intensities are denoted by $\mathbb{P}$ and $\mathbb{E}$, respectively.

The investment portfolio of the insurance company bears interest with intensity $r(t)$ at time t.

The intensities r and μ_{jk} constitute the *experience basis*, also called the *second order basis*, representing the true mechanisms governing the insurance business. At any time its past history is known, whereas its future is unknown.

We extend the set-up by viewing the second order basis as stochastic, whereby the uncertainty associated with it becomes quantifiable in probabilistic terms. In particular, prediction of its future development becomes a matter of model-based forecasting. Thus, let us consider the set-up above as the conditional model, given the second order basis, and place a distribution on the latter, whereby r and the μ_{jk} become stochastic processes. Let $\mathcal{G}_t$ denote their complete history up to, and including, time t and, accordingly, let $\mathbb{E}[\,\cdot\,|\,\mathcal{G}_t]$ denote conditional expectation, given this information.

For the time being we will work only in the conditional model and need not specify any particular marginal distribution of the second order elements.

B. The first order model. We let the first order model be of the same type as the conditional model of second order. Thus, the first order basis is viewed as deterministic, and we denote its elements by r^* and μ_{jk}^* and the corresponding probability measure and expectation operator by $\mathbb{P}^*$ and $\mathbb{E}^*$, respectively. The first order basis represents a prudent initial assessment of the development of the second order basis, and its elements are placed on the safe side in a sense that will be made precise later.

By statute, the insurer must currently provide a reserve to meet future liabilities in respect of the contract, and these liabilities are to be valuated on the first order basis. The *first order reserve* at time t, given that the policy is then in state j, is

$$\begin{aligned} V_j^*(t) &= \mathbb{E}^*\left[\int_t^n e^{-\int_t^\tau r^*} dB(\tau) \,\middle|\, Z(t)=j\right] \\ &= \int_t^n e^{-\int_t^\tau r^*} \sum_g p_{jg}^*(t,\tau)\left(dB_g(\tau) + \sum_{h;h\neq g} b_{gh}(\tau)\mu_{gh}^*(\tau)\,d\tau\right). \end{aligned} \tag{9.1}$$

We need Thiele's differential equations

$$dV_j^*(t) = r^*(t)V_j^*(t)\,dt - dB_j(t) - \sum_{k;\,k\neq j} R_{jk}^*(t)\,\mu_{jk}^*(t)\,dt\,, \tag{9.2}$$

where

$$R_{jk}^*(t) = b_{jk}(t) + V_k^*(t) - V_j^*(t) \tag{9.3}$$

is the *sum at risk* associated with a possible transition from state j to state k at time t.

The premiums are based on the *principle of equivalence* exercised on the first order valuation basis,

$$\mathbb{E}^*\left[\int_{0-}^n e^{-\int_0^\tau r^*}\,dB(\tau)\right] = 0\,, \tag{9.4}$$

or, equivalently,

$$V_0^*(0) = -\Delta B_0(0)\,. \tag{9.5}$$

9.3 The technical surplus and how it emerges

A. Definition of the mean portfolio surplus. With premiums determined by the principle of equivalence (9.4) based on prudent first order assumptions, the portfolio will create a systematic technical surplus if everything goes well. Quite naturally, the surplus is some average of past net incomes valuated on the factual second order basis less future net outgoes valuated on the conservative first order basis. The portfolio-wide mean surplus thus construed is

$$\begin{aligned} S(t) &= \mathbb{E}\left[\int_{0-}^t e^{\int_\tau^t r}\,d(-B)(\tau)\,\middle|\,\mathcal{G}_t\right] - \sum_j p_{0j}(0,t)\,V_j^*(t) \\ &= -e^{\int_0^t r}\int_{0-}^t e^{-\int_0^\tau r}\sum_j p_{0j}(0,\tau)\left(dB_j(\tau) + \sum_{k;k\neq j} b_{jk}(\tau)\mu_{jk}(\tau)\,d\tau\right) \\ &\quad - \sum_j p_{0j}(0,t)\,V_j^*(t)\,. \end{aligned} \tag{9.6}$$

The definition conforms with basic principles of insurance accountancy; at any time the balance is the difference between, on the debit, the factual income in the past and, on the credit, the reserve that by statute is to be provided in respect of future liabilities. In particular, due to (9.5),

$$S(0) = 0 \tag{9.7}$$

and, due to $V_j^*(n) = 0$,

$$S(n) = \mathbb{E}\left[\int_{0-}^{n} e^{\int_\tau^n r}\, d(-B)(\tau) \,\middle|\, \mathcal{G}_n\right], \tag{9.8}$$

as it ought to be.

Note that the expression in (9.6) involves only the past history of the second order basis, which is currently known.

B. The contributions to the surplus. Differentiating (9.6), applying the Kolmogorov forward equation (7.20) and the Thiele backward equation (9.2) to the last term on the right, leads to

$$\begin{aligned} dS(t) \;=\; & -e^{\int_0^t r}\, r(t)\, dt \int_{0-}^{t} e^{-\int_0^\tau r} \sum_j p_{0j}(0,\tau) \left(dB_j(\tau) + \sum_{k;k\neq j} b_{jk}(\tau)\mu_{jk}(\tau)\, d\tau \right) \\ & - \sum_j p_{0j}(0,t) \left(dB_j(t) + \sum_{k;k\neq j} b_{jk}(t)\mu_{jk}(t)\, dt \right) \\ & - \sum_j \left(\sum_{g;\, g\neq j} p_{0g}(0,t)\, \mu_{gj}(t)\, dt \;-\; p_{0j}(0,t)\, \mu_{j\cdot}(t)\, dt\,. \right) V_j^*(t) \\ & - \sum_j p_{0j}(0,t) \left(r^*(t) V_j^*(t)\, dt - dB_j(t) - \sum_{k;\, k\neq j} R_{jk}^*(t)\, \mu_{jk}^*(t)\, dt \right). \end{aligned}$$

Reusing the relation (9.6) in the first line here and gathering terms, we obtain

$$dS(t) = r(t)\, dt\, S(t) \;+\; \sum_j p_{0j}(0,t) c_j(t)\, dt\,,$$

with

$$c_j(t) = \{r(t) - r^*(t)\}\, V_j^*(t) + \sum_{k;\, k\neq j} R_{jk}^*(t) \{\mu_{jk}^*(t) - \mu_{jk}(t)\}\,. \tag{9.9}$$

Finally, integrating up and using (9.7), we arrive at

$$S(t) = \int_0^t e^{\int_\tau^t r} \sum_j p_{0j}(0,\tau) c_j(\tau)\, d\tau\,, \tag{9.10}$$

which expresses the technical surplus at any time as the sum of past contributions compounded with second order interest.

One may arrive at the definition of the contributions (9.9) by another route, starting from the *individual surplus* defined, quite naturally, as

$$S_{ind}(t) = e^{\int_0^t r} \int_{0-}^t e^{-\int_0^\tau r}\, d(-B)(\tau) - \sum_j I_j^Z(t) V_j^*(t). \tag{9.11}$$

Upon differentiating this expression, and proceeding along the same lines as above, one finds that $S_{ind}(t)$ consists of a purely erratic term and a systematic term. The latter is $\int_0^t e^{\int_\tau^t r} \sum_j I_j^Z(\tau) c_j(\tau)\, d\tau$, which is the individual counterpart of (9.10), showing how the contributions emerge at the level of the individual policy. They form a random payment function C defined by

$$dC(t) = \sum_j I_j^Z(t)\, c_j(t)\, dt\,. \tag{9.12}$$

With this definition, we can recast (9.10) as

$$S(t) = \mathbb{E}\left[\int_0^t e^{\int_\tau^t r} dC(\tau) \,\middle|\, \mathcal{G}_t \right]. \tag{9.13}$$

C. Safety margins. The expression on the right of (9.9) displays how the contributions arise from *safety margins* in the first order force of interest (the first term) and in the transition intensities (the second term). The purpose of the first order basis is to create a non-negative technical surplus. This is certainly fulfilled if

$$r(t) \geq r^*(t) \tag{9.14}$$

(assuming that all $V_j^*(t)$ are non-negative as they should be) and

$$\operatorname{sign}\{\mu_{jk}^*(t) - \mu_{jk}(t)\} = \operatorname{sign} R_{jk}^*(t)\,. \tag{9.15}$$

9.4 Dividends and bonus

A. The dividend process. Legislation lays down that the technical surplus belongs to the insured and has to be repaid in its entirety. Therefore, to the contractual payments B there must be added dividends, henceforth denoted by D. The dividends are currently adapted to the development of the second order basis and, as explained in Paragraph 9.1.A, they can not be negative. The purpose of the dividends is to establish, ultimately, equivalence on the true second order basis:

$$\mathbb{E}\left[\int_{0-}^n e^{-\int_0^\tau r}\, d\{B + D\}(\tau) \,\middle|\, \mathcal{G}_n \right] = 0\,. \tag{9.16}$$

We can state (9.16) equivalently as

$$\mathbb{E}\left[\int_{0-}^{n} e^{\int_\tau^n r}\, d\{B+D\}(\tau)\,\middle|\, \mathcal{G}_n\right] = 0\,. \tag{9.17}$$

The value at time t of past individual contributions less dividends, compounded with interest, is

$$U^d(t) = \int_{0-}^{t} e^{\int_\tau^t r}\, d\{C-D\}(\tau)\,. \tag{9.18}$$

This amount is an outstanding account of the insured against the insurer, and we shall call it the *dividend reserve* at time t.

By virtue of (9.8) and (9.13) we can recast the equivalence requirement (9.17) in the appealing form

$$\mathbb{E}[U^d(n)\,|\,\mathcal{G}_n] = 0\,. \tag{9.19}$$

From a solvency point of view it would make sense to strengthen (9.19) by requiring that compounded dividends must never exceed compounded contributions:

$$\mathbb{E}[U^d(t)\,|\,\mathcal{G}_t] \geq 0\,, \tag{9.20}$$

$t \in [0, n]$. At this point some explanation is in order. Although the ultimate balance requirement is enforced by law, the dividends do not represent a *contractual* obligation on the part of the insurer; the dividends must be adapted to the second order development up to time n and can, therefore, not be stipulated in the terms of the contract at time 0. On the other hand, at any time, dividends allotted in the past have irrevocably been credited to the insured's account. These regulatory facts are reflected in (9.20).

If we adopt the view that "the technical surplus belongs to those who created it", we should sharpen (9.19) by imposing the stronger requirement

$$U^d(n) = 0\,. \tag{9.21}$$

This means that no transfer of redistributions across policies is allowed. The solvency requirement conforming with this point of view, and sharpening (9.20), is

$$U^d(t) \geq 0\,, \tag{9.22}$$

$t \in [0, n]$.

The constraints imposed on D in this paragraph are of a general nature and leave a certain latitude for various designs of dividend schemes. We shall list some possibilities motivated by practice.

B. Special dividend schemes. The so-called *contribution scheme* is defined by $D = C$, that is, all contributions are currently and immediately credited to the account of the insured. No dividend reserve will accrue and, consequently, the only instrument on the part of the insurer in case of adverse second order experience is to cease crediting dividends. In some countries the contribution principle is enforced by law. This means that insurers are compelled to operate with minimal protection against adverse second order developments.

By *terminal dividend* is meant that all contributions are currently invested and their compounded total is credited to the insured as a lump sum dividend payment only upon the termination of the contract at some time T after which no more contributions are generated. Typically T would be the time of transition to an absorbing state (death or withdrawal), truncated at n. If compounding is at second order rate of interest, then

$$D(t) = 1[t \geq T] \int_0^T e^{\int_\tau^T r} dC(\tau) .$$

Contribution dividends and terminal dividends represent opposite extremes in the set of conceivable dividend schemes, which are countless. One class of intermediate solutions are those that yield dividends only at certain times $T_1 < \cdots < T_K \leq n$, e.g. annually or at times of transition between certain states. At each time T_i the amount $\Delta D(T_i) = \int_{T_{i-1}}^{T_i} e^{\int_\tau^{T_i} r} dC(\tau)$ (with $T_0 = 0$) is entered to the insured's credit.

C. Allocation of dividends; bonus. Once they have been allotted, dividends belong to the insured. They may, however, be disposed of in various ways and need not be paid out currently as they fall due. The actual payouts of dividends are termed *bonus* in the sequel, and the corresponding payment function is denoted by B^b.

The compounded value of credited dividends less paid bonuses at time t is

$$U^b(t) = \int_0^t e^{\int_\tau^t r} d\{D - B^b\}(\tau) . \tag{9.23}$$

This is a debt owed by the insurer to the insured, and we shall call it the *bonus reserve* at time t. Bonuses may not be advanced, so B^b must satisfy

$$U^b(t) \geq 0 \tag{9.24}$$

for all $t \in [0, n]$. In particular, since $D(0) = 0$, one has $B^b(0) = 0$. Moreover, since all dividends must eventually be paid out, we must have

$$U^b(n) = 0 . \tag{9.25}$$

We have introduced three notions of reserves that all appear on the debit side of the insurer's balance sheet. First, the premium reserve V^* is provided to meet net outgoes in respect of future events; second, the dividend reserve U^d is provided to settle the excess of past contributions over past dividends; third, the bonus reserve U^b is provided to settle the unpaid part of dividends credited in the past. The premium reserve is of prospective type and is a predicted amount, whereas the dividend and bonus reserves are of retrospective type and are indeed known amounts summing up to

$$U^d(t) + U^b(t) = \int_0^t e^{\int_\tau^t r}\, d\{C - B^b\}(\tau)\,, \tag{9.26}$$

the compounded total of past contributions not yet paid back to the insured.

D. Some commonly used bonus schemes. The term *cash bonus* is, quite naturally, used for the scheme $B^b = D$. Under this scheme the bonus reserve is always null, of course.

By *terminal bonus*, also called *reversionary bonus*, is meant that all dividends, with accumulation of interest, are paid out as a lump sum upon the termination of the contract at some time T, that is,

$$B^b(t) = 1[t \geq T] \int_0^T e^{\int_\tau^T r} dD(\tau)\,.$$

Here we could replace the integrator D by C since terminal bonus obviously does not depend on the dividend scheme; all contributions are to be repaid with accumulation of interest.

Assume now, what is common in practice, that dividends are currently used to purchase additional insurance coverage of the same type as in the primary policy. It seems natural to let the *additional benefits* be proportional to those stipulated in the primary policy since they represent the desired profile of the product. Thus, the dividends $dD(s)$ in any time interval $[s, s+ds)$ are used as a single premium for an insurance with payment function of the form

$$dQ(s)\{B^+(\tau) - B^+(s)\}\,,$$

$\tau \in (s, n]$, where the topscript "$+$" signifies, in an obvious sense, that only positive payments (benefits) are counted.

Supposing that additional insurances are written on first order basis, the proportionality factor $dQ(s)$ is determined by

$$dD(s) = dQ(s) V^{*+}_{Z(s)}(s), \tag{9.27}$$

where

$$V^{*+}_{Z(s)}(s) = \mathbb{E}^* \left[\int_s^n e^{-\int_s^\tau r^*} dB^+(\tau) \middle| \, Z(s) \right]$$

is the single premium at time s for the future benefits under the policy.

Now the bonus payments B^b are of the form

$$dB^b(t) = Q(t) dB^+(t)\,. \tag{9.28}$$

Being written on first order basis, also the additional insurances create technical surplus. The total contributions under this scheme develop as

$$dC(t) + Q(t) dC^+(t)\,, \tag{9.29}$$

where the first term on the right stems from the primary policy and the second term stems from the $Q(t)$ units of additional insurances purchased in the past, each of which has payment function B^+ producing contributions C^+ of the form $dC^+(t) = \sum_j I_j^Z(t)\, c_j^+(t)\, dt$, with

$$c_j^+(t) = \{r(t) - r^*(t)\} V_j^{*+}(t) + \sum_{k;\, k \neq j} R_{jk}^{*+}(t) \{\mu_{jk}^*(t) - \mu_{jk}(t)\} ,$$

$$R_{jk}^{*+}(t) = b_{jk}^+(t) + V_k^{*+}(t) - V_j^{*+}(t) .$$

The present situation is more involved than those encountered previously since, not only are dividends driven by the contractual payments, but it is also the other way around. To keep things relatively simple, suppose that the contribution principle is adopted so that the dividends in (9.27) are set equal to the contributions in (9.29). Then the system is governed by the dynamics

$$dC(t) + Q(t) dC^+(t) = dQ(t) V_{Z(t)}^{*+}(t)$$

or, realizing that $V_{Z(t)}^{*+}(t)$ is strictly positive whenever $dC(t)$ and $dC^+(t)$ are,

$$dQ(t) - Q(t)\, dG(t) = dH(t) , \tag{9.30}$$

where G and H are defined by

$$dG(t) = \frac{1}{V_{Z(t)}^{*+}(t)}\, dC^+(t) , \tag{9.31}$$

$$dH(t) = \frac{1}{V_{Z(t)}^{*+}(t)}\, dC(t) . \tag{9.32}$$

Multiplying with $\exp(-G(t))$ to form a complete differential on the left and then integrating from 0 to t, using $Q(0) = 0$, we obtain

$$Q(t) = \int_0^t e^{G(t) - G(\tau)}\, dH(\tau) . \tag{9.33}$$

9.5 Bonus prognoses

A. A Markov chain environment. We shall adopt a simple Markov chain description of the uncertainty associated with the development of the second order basis. Let $Y(t)$, $0 \leq t \leq n$, be a time-continuous Markov chain with finite state space $\mathcal{Y} = \{1, \ldots, q\}$ and constant intensities of transition, λ_{ef}. Denote the associated indicator processes by I_e^Y. The process Y represents the "economic-demographic environment", and we let the second order elements depend on the current Y-state:

$$r(t) = \sum_e I_e^Y(t)\, r_e = r_{Y(t)} ,$$

$$\mu_{jk}(t) = \sum_e I_e^Y(t)\, \mu_{e;jk}(t) = \mu_{Y(t);jk}(t) .$$

The r_e are constants and the $\mu_{e;jk}(t)$ are intensity functions, all deterministic.

With this specification of the full two-stage model it is realized that the pair $X = (Y, Z)$ is a Markov chain on the state space $\mathcal{X} = \mathcal{Y} \times \mathcal{Z}$, and its intensities of transition, which we denote by $\kappa_{ej,fk}(t)$ for (e, j), $(f, k) \in \mathcal{X}$, $(e, j) \neq (f, k)$, are

$$\kappa_{ej,fj}(t) = \lambda_{ef}\,, \qquad e \neq f, \tag{9.34}$$
$$\kappa_{ej,ek}(t) = \mu_{e;jk}(t)\,, \quad j \neq k, \tag{9.35}$$

and null for all other transitions.

In this extended set-up the contributions, whose dependence on the second order elements was not visualized earlier, can appropriately be represented as

$$dC(t) = c(t)\,dt = \sum_{e,j} I_e^Y(t) I_j^Z(t) c_{ej}(t)\,dt,$$

where

$$c_{ej}(t) = \{r_e - r^*(t)\}\, V_j^*(t) + \sum_{k;\,k\neq j} R_{jk}^*(t)\{\mu_{jk}^*(t) - \mu_{e;jk}(t)\}\,. \tag{9.36}$$

Under the scheme of additional benefits described in Paragraph 9.4.D a similar convention goes for C^+ and c^+ and, accordingly, (9.31) and (9.32) become

$$dG(t) \;=\; g(t)\,dt = \sum_{e,j} I_e^Y(t) I_j^Z(t) g_{ej}(t)\,dt\,, \tag{9.37}$$
$$g_{ej}(t) \;=\; \frac{c_{ej}^+(t)}{V_j^{*+}(t)}\,, \tag{9.38}$$
$$dH(t) \;=\; h(t)\,dt = \sum_{e,j} I_e^Y(t) I_j^Z(t) h_{ej}(t)\,dt\,, \tag{9.39}$$
$$h_{ej}(t) \;=\; \frac{c_{ej}(t)}{V_j^{*+}(t)}\,. \tag{9.40}$$

B. Preparatory remarks on the issue of bonus prognoses. There is no single functional of the future bonus stream that presents itself as *the* relevant quantity to prognosticate. One could e.g. take the total bonuses discounted by some suitable inflation rate, or the undiscounted total bonuses, or the rate at which bonus will be paid at certain times, and one could apply any of these possibilities to the random development of the policy or to some representative fixed development. We shall focus on the expected value, and in the simplest cases also higher order moments, of the future bonuses discounted by the stochastic second order interest. From this we can easily deduce predictors for a number of other relevant quantities. We turn now to the analysis of some of the schemes described in Section 9.4.

C. Contribution dividends and cash bonus. This case, where $B^b = C = D$, is particularly simple since the bonus payments at any time depend only on the current state of the process. We can then employ the appropriate version of Thiele's differential equation to calculate the state-wise expected discounted future bonuses (= contributions),

$$W_{ej}(t) = \mathbb{E}\left[\int_t^n e^{-\int_t^\tau r}\, c(\tau)\, d\tau \,\middle|\, X(t) = (e,j) \right] .$$

They are determined by the appropriate version of Thiele's differential equation,

$$\begin{aligned} \frac{d}{dt} W_{ej}(t) &= r_e\, W_{ej}(t) - c_{ej}(t) - \sum_{f;f\neq e} \lambda_{ef}\, (W_{fj}(t) - W_{ej}(t)) \\ &\quad - \sum_{k;k\neq j} \mu_{e;jk}(t)\, (W_{ek}(t) - W_{ej}(t)) \,, \end{aligned} \tag{9.41}$$

subject to

$$W_{ej}(n-) = 0\,, \quad \forall e, j\,. \tag{9.42}$$

D. Terminal dividend and/or bonus. Under the terminal bonus scheme dividends and bonuses are the same, of course. The problem of predicting the total bonus payments discounted with respect to second order interest is basically the same as in the previous paragraph since it amounts to adding the total amount of compounded past contributions, which is known, and the state-wise predictor of discounted future contributions.

Suppose instead that at time t, the policy still being in force, it is decided to predict the undiscounted value of the terminal bonus amount,

$$W = \int_0^T e^{\int_\tau^T r} c(\tau)\, d\tau = \int_0^t e^{\int_\tau^t r}\, c(\tau)\, d\tau\, W'(t) + W''(t)\,, \tag{9.43}$$

where

$$\begin{aligned} W'(t) &= e^{\int_t^T r}, \\ W''(t) &= \int_t^T e^{\int_\tau^T r} c(\tau)\, d\tau\,. \end{aligned}$$

We need the state-wise expected values

$$\begin{aligned} W'_e(t) &= \mathbb{E}[W'(t) \,|\, Y(t) = e], \\ W''_{ej}(t) &= \mathbb{E}[W''(t) \,|\, X(t) = (e,j)]\,, \end{aligned}$$

to find the state-wise predictors of W in (9.43),

$$W_{ej}(t) = \int_0^t e^{\int_\tau^t r}\, c(\tau)\, d\tau\, W'_e(t) + W''_{ej}(t)\,.$$

We shall find these functions by the backward construction, starting from

$$\begin{aligned} W'(t) &= e^{r\,dt}\, W'(t+dt), \\ W''(t) &= c(t)\,dt\, W'(t) + W''(t+dt)\,. \end{aligned}$$

Conditioning on what happens in the small time interval $(t, t+dt]$, we get

$$W'_e(t) = e^{r_e\,dt} \left((1-\lambda_{e\cdot}\,dt)\, W'_e(t+dt) + \sum_{f;\, f\neq e} \lambda_{ef}(t)\,dt\, W'_f(t+dt) \right),$$

and

$$\begin{aligned} W''_{ej}(t) &= c_{ej}(t)\,dt\,W'_e(t) + (1-(\lambda_{e\cdot} + \mu_{e;j\cdot}(t))\,dt)\,W''_{ej}(t+dt) \\ &\quad + \sum_{f;\, f\neq e} \lambda_{ef}(t)\,dt\,W''_{fj}(t+dt) \\ &\quad + \sum_{k;\, k\neq j} \mu_{e;jk}(t)\,dt\,W''_{ek}(t+dt)\,. \end{aligned}$$

From these relationships we easily obtain the differential equations

$$\frac{d}{dt}W'_e(t) = -r_e\,W'_e(t) - \sum_{f;\, f\neq e} \lambda_{ef}\,\left(W'_f(t) - W'_e(t)\right), \tag{9.44}$$

$$\begin{aligned} \frac{d}{dt}W''_{ej}(t) &= -\,c_{ej}(t)\,W'_e(t) - \sum_{f;\, f\neq e} \lambda_{ef}\,\left(W''_{fj}(t) - W''_{ej}(t)\right) \\ &\quad - \sum_{k;\, k\neq j} \mu_{e;jk}(t)\,\left(W''_{ek}(t) - W''_{ej}(t)\right), \end{aligned} \tag{9.45}$$

which are to be solved subject to

$$W'_e(n-) = 1\,, \qquad W''_{ej}(n-) = 0\,, \qquad \forall e, j\,. \tag{9.46}$$

E. Additional benefits. Suppose we want to predict the total future bonuses discounted with respect to second order interest,

$$W(t) = \int_t^n e^{-\int_t^\tau r}\, Q(\tau)\,dB^+(\tau)\,,$$

with Q defined by (9.33). Recalling (9.37)–(9.40), we reshape $W(t)$ as

$$\begin{aligned} W(t) &= \int_t^n e^{-\int_t^\tau r} \int_0^\tau e^{\int_r^\tau g}\, h(r)\,dr\,dB^+(\tau) \\ &= \int_t^n e^{-\int_t^\tau r} \left(\int_0^t e^{\int_r^t g}\, h(r)\,dr\, e^{\int_t^\tau g} + \int_t^\tau e^{\int_r^\tau g}\, h(r)\,dr \right) dB^+(\tau) \\ &= \int_0^t e^{\int_r^t g}\, h(r)\,dr\, W'(t) + W''(t)\,, \end{aligned} \tag{9.47}$$

with

$$
\begin{aligned}
W'(t) &= \int_t^n e^{\int_t^\tau (g-r)}\, dB^+(\tau), \\
W''(t) &= \int_t^n e^{-\int_t^\tau r}\, W'(\tau)\, h(\tau)\, d\tau\,.
\end{aligned}
$$

Thus, we need the state-wise expected values

$$
\begin{aligned}
W'_{ej}(t) &= \mathbb{E}[W'(t)\,|\,X(t)=(e,j)]\,, \\
W''_{ej}(t) &= \mathbb{E}[W''(t)\,|\,X(t)=(e,j)]\,,
\end{aligned}
$$

in order to find the state-wise predictors of $W(t)$ in (9.47),

$$
W_{ej}(t) = \int_0^t e^{\int_r^t g}\, h(r)\, dr\, W'_{ej}(t) + W''_{ej}(t)\,.
$$

The backward equations start from

$$
\begin{aligned}
W'(t) &= dB^+(t) + e^{(g(t)-r(t))\,dt}\, W'(t+dt)\,, \\
W''(t) &= W'(t)\, h(t)\, dt + e^{-r(t)\,dt}\, W''(t+dt)\,,
\end{aligned}
$$

from which we proceed in the same way as in the previous paragraph to obtain

$$
\begin{aligned}
dW'_{ej}(t) &= -\,dB_j^+(t) + (r_e - g_{ej}(t))\,dt\,W'_{ej}(t) \\
&\quad - \sum_{f;\,f\neq e} \lambda_{ef}\,dt\,\left(W'_{fj}(t) - W'_{ej}(t)\right) \\
&\quad - \sum_{k;k\neq j} \mu_{e;jk}(t)\,dt\,\left(b_{jk}^+(t) + W'_{ek}(t) - W'_{ej}(t)\right)\,, \qquad (9.48) \\
dW''_{ej}(t) &= -\,W'_{ej}(t)h_{ej}(t)\,dt + r_e\,dt\,W''_{ej}(t) \\
&\quad - \sum_{f;\,f\neq e} \lambda_{ef}\,dt\,\left(W''_{fj}(t) - W''_{ej}(t)\right) \\
&\quad - \sum_{k;\,k\neq j} \mu_{e;jk}(t)\,dt\,\left(W''_{ek}(t) - W''_{ej}(t)\right)\,. \qquad (9.49)
\end{aligned}
$$

The appropriate side conditions are

$$
W'_{ej}(n-) = \Delta B_j^+(n)\,, \qquad W''_{ej}(n-) = 0\,, \qquad \forall e,j\,. \qquad (9.50)
$$

F. Predicting undiscounted amounts. If the undiscounted total contributions or additional benefits is what one wants to predict, one can just apply the formulas with all r_e replaced by 0.

G. Predicting bonuses for a given policy path. Yet another form of prognosis, which may be considered more informative than the two mentioned above, would be to predict bonus payments for some possible fixed pursuits of a policy instead of averaging over all possibilities. Such prognoses are obtained from those described above upon keeping the realized path $Z(\tau)$ for $\tau \in [0,t]$, where t is the time of consideration, and putting $Z(\tau) = z(\tau)$ for $\tau \in (t,n]$, where $z(\cdot)$ is some fixed path with $z(t) = Z(t)$. The relevant predictors then become essentially functions only of the current Y-state and are simple special cases of the results above.

As an example of an even simpler type of prognosis for a policy in state j at time t, the insurer could present the expected bonus payment per time unit at a future time s, given that the policy is then in state i, and do this for some representative selections of s and i. If $Y(t) = e$, then the relevant prediction is

$$\mathbb{E}[c_{Y(s)i}(s) \,|\, Y(t) = e] = \sum_f p^Y_{ef}(t,s) c_{fi}(s) \,.$$

9.6 Examples

A. The case. For our purpose, which is to illustrate the role of the stochastic environment in model-based prognoses, it suffices to consider simple insurance products for which the relevant policy states are $\mathcal{Z} = \{a, d\}$ ('alive' and 'dead').

We will consider a single life insured at age 30 for a period of $n = 30$ years, and let the first order elements be those of the Danish technical basis G82M for males:

$$\begin{aligned} r^* &= \ln(1.045)\,, \\ \mu^*_{ad}(t) &= \mu^*(t) = 0.0005 + 0.000075858 \cdot 10^{0.038(30+t)}\,. \end{aligned}$$

Three different forms of insurance benefits will be considered, and in each case we assume that premiums are payable continuously at level rate as long as the policy is in force. First, a term insurance (TI) of $1 = b_{ad}(t)$ with first order premium rate $0.0042608 = -b_a(t)$. Second, a pure endowment (PE) of $1 = \Delta B_a(30)$ with first order premium rate $0.0140690 = -b_a(t)$. Third, an endowment insurance (EI), which is just the combination of the former two; $1 = b_{ad}(t) = \Delta B_a(30)$, $0.0183298 = -b_a(t)$.

Just as an illustration, let the second order model be the simple one where interest and mortality are governed by independent time-continuous Markov chains and, more specifically, that r switches with a constant intensity λ_i between the first order rate r^* and a better rate $\epsilon_i r^*$ ($\epsilon_i > 1$) and, similarly, μ switches with a constant intensity λ_m between the first order rate μ^* and a better rate $\epsilon_m \mu^*$ ($\epsilon_m < 1$). (We choose to express ourselves this way although (9.15) shows that, for insurance forms with negative sum at risk, e.g. pure endowment insurance, it is actually a higher second order mortality that is "better" in the sense of creating positive contributions.)

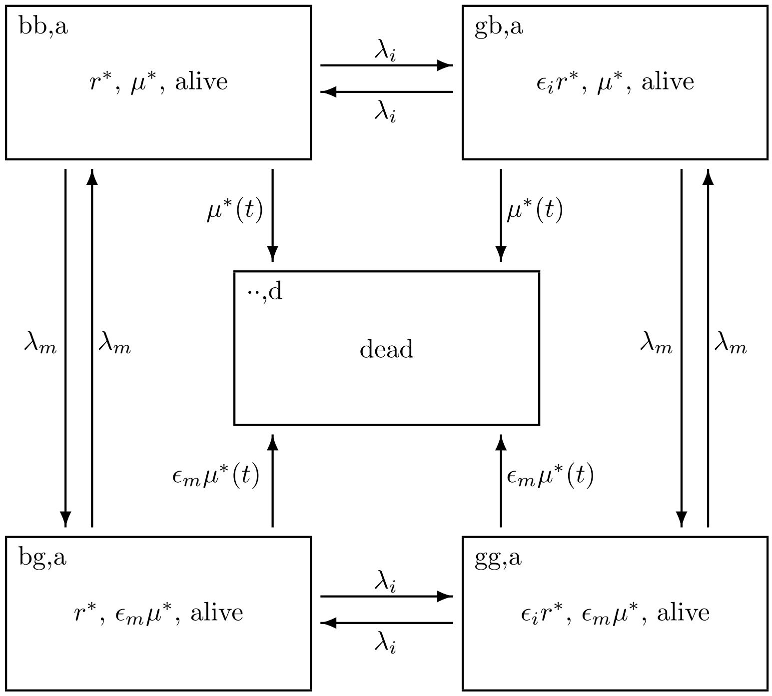


Figure 9.1: The Markov process $X = (Y, Z)$ for a single life insurance in an environment with two interest states and two mortality states.

The situation fits into the framework of Paragraph 9.5.A; Y has states $\mathcal{Y} = \{bb, gb, bg, gg\}$ representing all combinations of "bad" (b) and "good" (g) interest and mortality, and the non-null intensities are

$$\lambda_{bb,gb} = \lambda_{gb,bb} = \lambda_{bg,gg} = \lambda_{gg,bg} = \lambda_i \,,$$

$$\lambda_{bb,bg} = \lambda_{bg,bb} = \lambda_{gb,gg} = \lambda_{gg,gb} = \lambda_m \,.$$

The first order basis is just the worst-scenario bb.

Adopting the device (9.34)–(9.35), we consider the Markov chain $X = (Y, Z)$ with states (bb, a), (gb, a), etc. It is realized that all death states can be merged into one, so it suffices to work with the simple Markov model with five states sketched in Figure 9.1.

B. Results. We shall report some numerical results for the case where $\epsilon_i = 1.25$, $\epsilon_m = 0.75$, and $\lambda_i = \lambda_m = 0.1$. Prognoses are made at the time of issue of the policy. Computations were performed by the fourth order Runge-Kutta method, which turns out to work with high precision in the present class of situations.

Table 9.1 displays, for each of the three policies, the state-wise expected values of discounted contributions obtained by solving (9.41)–(9.42). We shall be content here to point out two features: First, for the term insurance the mortality margin is far more important than the interest margin, whereas for the pure endowment it is the other way around (the latter has the larger reserve). Note that the sum at risk is negative for the pure endowment, so that the first order assumption of excess mortality is really not to the safe side, see (9.15). Second, high interest produces large contributions, but, since high initial interest also induces severe discounting, it is not necessarily true that good initial interest will produce a high value of the expected discounted contributions, see the two last entries in the row TI.

The latter remark suggests the use of a discounting function different from the one based on the second order interest, e.g. some exogenous deflator reflecting the likely development of the price index or the discounting function corresponding to first order interest. In particular, one can simply drop discounting and prognosticate the total amounts paid. We shall do this in the following, noting that the expected value of bonuses discounted by second order interest must in fact be the same for all bonus schemes, and are already shown in Table 9.1.

Table 9.2 shows state-wise expected values of undiscounted bonuses for three different schemes; contribution dividends and cash bonus (C, the same as total undiscounted contributions), terminal bonus (TB), and additional benefits (AB).

We first note that, now, any improvement of initial second order conditions helps to increase prospective contributions and bonuses.

Furthermore, expected bonuses are generally smaller for C than for TB and AB since bonuses under C are paid earlier. Differences between TB and AB must be due to a similar effect. Thus, we can infer that AB must on the average fall due earlier than TB, except for the pure endowment policy, of course.

One might expect that the bonuses for the term insurance and the pure endowment policies add up to the bonuses for the combined endowment insurance policy, as is the case for C and TB. However, for AB it is seen that the sum of the bonuses for the two component policies is generally smaller than the bonuses for the combined policy. The explanation must be that additional death benefits and additional survival benefits are not purchased in the same proportions under the two policy strategies. The observed difference indicates that, on the average, the additional benefits fall due later under the combined policy, which therefore must have the smaller proportion of additional death benefits.

C. Assessment of prognostication error. Bonus prognoses based on the present model may be equipped with quantitative measures of the prognostication error. By the technique of proof shown in Section 9.5 we may derive differential equations for higher order moments of any of the predictands considered and calculate e.g. the coefficient of variation, the skewness, and the kurtosis.

Table 9.1: Conditional expected present value at time 0 of total contributions for term insurance policy (TI), pure endowment policy (PE), and endowment insurance policy (EI), given initial second order states of interest and mortality (b or g).

	bb	gb	bg	gg
TI :	.00851	.00854	.01061	.01059
PE :	.01613	.01823	.01595	.01807
EI :	.02463	.02677	.02656	.02865

9.7 Including expenses

A. The form of the expenses. Expenses are assumed to incur in accordance with a non-decreasing payment function A of the same type as the contractual payments, that is,

$$dA(t) = \sum_j I_j^Z(t)\, dA_j(t) + \sum_{j \neq k} a_{jk}(t)\, dN_{jk}^Z(t)\,. \tag{9.51}$$

It is common in practice to assume, furthermore, that expenses of annuity type incur with a lump sum of initial costs at time 0 and thereafter continuously at a rate that depends on the current state, that is, $\Delta A_0(0) > 0$ and $dA_j(t) = a_j(t)\, dt$ for $t > 0$. The transition costs $a_{jk}(t)$ are not explicitly taken into account in practice, but we include them here since they add realism without adding mathematical complexity.

B. First order assumptions. The elements $\Delta A_0(0)$, $a_j(t)$, and $a_{jk}(t)$ will in general depend on the second order development, and the first order basis must, therefore, specify prudent estimates $\Delta A_0^*(0)$, $a_j^*(t)$, and $a_{jk}^*(t)$. Denote the corresponding payment function by A^*.

C. Surplus and contributions in the presence of expenses. The introduction of expenses adds a new feature to the previous set-up in that also the payments become dependent on the second order development. However, the essential parts of the analyses in the previous sections carry over with merely notational modifications; all it takes is to replace everywhere the contractual payment function B with $A + B$ in the past and $A^* + B$ in the future. One finds, in particular, that the first order equivalence relation (9.5) now turns into

$$V_0^*(0) = -\Delta A_0^*(0) - \Delta B_0(0)\,, \tag{9.52}$$

the surplus at time 0 becomes

$$S(0) = \Delta A_0^*(0) - \Delta A_0(0)\,, \tag{9.53}$$

Table 9.2: Conditional expected value (E) of undiscounted total contributions (C), terminal bonus (TB), and total additional benefits (AB) for term insurance policy (TI), pure endowment policy (PE), and endowment insurance policy (EI), given initial second order states of interest and mortality (b or g).

			bb	gb	bg	gg
TI:	E	C :	.02153	.02222	.02436	.02505
	E	TB :	.03693	.03916	.04600	.04847
	E	AB :	.02949	.03096	.03545	.03706
PE:	E	C :	.04342	.04818	.04314	.04791
	E	TB :	.07337	.08687	.07264	.08615
	E	AB :	.07337	.08687	.07264	.08615
EI:	E	C :	.06495	.07040	.06750	.07296
	E	TB :	.11030	.12603	.11864	.13462
	E	AB :	.10723	.12199	.11501	.13003

and the contributions consist of a jump

$$\Delta C(0) = \Delta A_0^*(0) - \Delta A_0(0) \tag{9.54}$$

at time 0 and thereafter a continuous part, which is defined upon replacing (9.9) with

$$\begin{aligned} c_j(t) &= \{r(t) - r^*(t)\}\, V_j^*(t) + \{a_j^*(t) - a_j(t)\} \\ &\quad + \sum_{k;\, k\neq j} \{a_{jk}^*(t) - a_{jk}(t)\}\mu_{jk}(t) \\ &\quad + \sum_{k;\, k\neq j} R_{jk}^*(t)\{\mu_{jk}^*(t) - \mu_{jk}(t)\}\,, \end{aligned} \tag{9.55}$$

where now

$$R_{jk}^*(t) = a_{jk}^*(t) + b_{jk}(t) + V_k^*(t) - V_j^*(t)\,. \tag{9.56}$$

Referring to the discussion in Paragraph 9.3.C, we see that the contributions emerge from safety margins in all first order elements, r^*, μ_{jk}^*, and A^*.

D. Prediction in the presence of expenses. The complexity of the prediction problem depends heavily on the assumptions made about the second order expenses, and at this point some new problems may arise.

Just to get started, suppose first that the expense elements $\Delta A_0(0)$, $a_j(t)$, and $a_{jk}(t)$ are deterministic. Then the methods in Section 9.5 carry over with only trivial modifications. Presumably, this simplistic model is at the base of the frequently encountered claim that "administration expenses can be regarded as additional benefits". Unfortunately, real life expenses are of a different, and typically less pleasant, nature. An exhaustive discussion of this issue could easily exhaust the reader, so we shall be content with just outlining some tentative ideas.

The problem is that expenses are made up of wages, commissions, rent, taxes and other items that are governed by the economic development. In the framework of the Markov model in Paragraph 9.5.A, one simple way of accounting for such effects is to make the second order expenses dependent on the current state of Y, that is,

$$\begin{aligned} \Delta A_0(0) &= \sum_e I_e^Y(t)\, \Delta A_{e0}(0)\,, \\ a_j(t) &= \sum_e I_e^Y(t)\, a_{ej}(t)\,, \\ a_{jk}(t) &= \sum_e I_e^Y(t)\, a_{ejk}(t)\,, \end{aligned}$$

with deterministic ΔA_{e0}, a_{ej}, and a_{ejk}. By enriching sufficiently the state space of Y, one can in principle create a fairly realistic model.

Perhaps the most reasonable point of view is that expenses are inflated by some time-dependent rate $\gamma(t)$ so that we should put $a_j(t) = e^{\int_0^t \gamma} a_j^0(t)$ and $a_{jk}(t) = e^{\int_0^t \gamma} a_{jk}^0(t)$ with a_j^0 and a_{jk}^0 deterministic. One possibility is to put the second order force of interest r in the role of γ. More realistically one should let γ be something else, but still related to r through joint dependence on a suitably specified Y. We shall not pursue this idea any further here, but note, by way of warning, that prognostication in this kind of inflation model will present problems in addition to those solved in Section 9.5.

9.8 Discussions

A. The principle of equivalence. This principle, as formulated in (9.4), is basic in life insurance. The expected value represents averaging over a large (really infinite) portfolio of policies, the philosophy being that, even if the individual policy creates a (possibly large) loss or gain, there will be balance on the average between outgoes and incomes in the portfolio as a whole if the premiums are set by equivalence. The deviation from perfect balance, which is inevitable in a finite world with finite portfolios, represents profit or loss on the part of the insurer and has to be settled by an adjustment of the equity capital. (The possibility of loss, about as likely and about as large as the possible profit, might seem unacceptable to an industry that needs to attract investors, but it should be kept in mind that salaries to employees and dividends to owners are accounted as part of the expenses discussed in Section 9.7.)

B. On the notion of second order basis. The definition of the second order basis as the true one is slightly at variance with practical usage (which is not uniform anyway). The various amendments made to our idealized definition in practice are due to administrative and procedural bottlenecks: The factual development of interest, mortality, etc. has to be verified by the insurer and then approved by the supervisory authority. Since this can not be a continuous operation, any regulatory definition of the second order basis must to some extent involve realistic, still typically conservative, short term forecasts of the future development. However, our definition can certainly be agreed upon as the intended one.

C. Model deliberations. The Markov chain model is mathematically tractable since state-wise expected values are determined by solving (in most cases simple) systems of first order ordinary differential equations. At the same time, when equipped with a sufficiently rich state space and appropriate intensities of transition, it is able to picture virtually any conceivable notion of the real object of the model.

The Markov chain model is particularly apt to describe the development of life insurance policy since the paths of Z are of the same kind as the true ones.

When used to describe the development of the second order basis, however, the approximative nature of the Markov chain is obvious, and it will surface immediately as e.g. the experienced force of interest takes values outside of the finite set allowed by the model. This is not a serious objection, however, and the next paragraph explains why.

D. The role of the stochastic environment model. A paramount concern is that of establishing equivalence conditional on the factual second order history in the sense of (9.16). Now, in this conditional expectation the marginal distribution of the second order elements does not appear and is, in this respect, irrelevant. Also the contributions and, hence, the dividends are functions only of the realized experience basis and do not involve the distribution of its elements.

Then, what remains the purpose of placing a distribution on the second order elements is to form a basis for prognostication of bonus. Subsidiary as it is, this role is still an important part of the play; although a prognosis does not commit the insurer to pay the forecasted amounts, it should as much as possible be a reliable piece of information to the insured. Therefore, the distribution placed on the second order elements should set a reasonable scenario for the course of events, but it need not be perfectly true. This is comforting since any view of the mechanisms governing the economic-demographic development is to some extent guess-work. When the accounts are eventually made up, every speculative element must be absent, and that is precisely what the principle (9.16) lays down.

E. A digression: Which is more important, interest or mortality? Actuarial wisdom says it is interest. This is, of course, an empirical statement based on the fact that, in the era of contemporary insurance, mortality rates have been smaller and more stable than interest rates. Our model can add some other kind of insight. We shall again be content with a simple illustration related to the single life described in Section 9.6. Table 9.3 displays expected values and standard deviations of the present values at time 0 of a term life insurance and a life annuity under various scenarios with fixed interest and mortality, that is, conditional on fixed Y-state throughout the term of the policy. The impact of interest variation is seen by reading column-wise, and the impact of mortality variation is seen by reading row-wise. The overall impression is that mortality is the more important element by term insurance, whereas interest is the (by far) more important by life annuity insurance.

Table 9.3: Expected value (E) and standard deviation (SD) of present values of a term life insurance (TI) with sum 1 and a life annuity (LA) with level intensity 1 per year, with interest $r = \epsilon_i r^*$ and mortality $\mu = \epsilon_m \mu^*$ for various choices of ϵ_i and ϵ_m.

			TI			LA		
		ϵ_m :	1.5	1.0	0.5	1.5	1.0	0.5
ϵ_i :	0.5	$\mathbb{E}$:	.14636	.10119	.05250	20.545	20.996	21.467
		SD :	.27902	.24104	.18041	03.691	03.101	02.257
	1.0	E :	.09927	.06834	.03531	15.750	16.039	16.340
		SD :	.20245	.17330	.12857	02.505	02.097	01.521
	1.5	E :	.06976	.04782	.02460	12.466	12.655	12.852
		SD :	.15858	.13437	.09868	01.759	01.468	01.061

Chapter 10

Miscellaneous topics

10.1 Woolhouse's formula

Let f be a real-valued differentiable function defined on $[0,1]$. We approximate f by a trinomial $a_0 + a_1 t + a_2 t^2 + a_3 t^3$ chosen such that its value and its slope coincide with those of f at the endpoints 0 and 1 of the interval:

$$\begin{aligned} a_0 &= f(0) \\ a_1 &= f'(0) \\ a_0 + a_1 + a_2 + a_3 &= f(1) \\ a_1 + 2a_2 + 3a_3 &= f'(1) \end{aligned}$$

The coefficients a_0 and a_1 are given explicitly by the first two equations, and we easily solve the last two equations to obtain $a_2 = -3f(0)+3f(1)-2f'(0)-f'(1)$ and $a_3 = 2f(0) - 2f(1) + f'(0) + f'(1)$. The integral of the trinomial gives an approximation to the integral of f:

$$\int_0^1 f(t)\,dt \approx a_0 + \frac{1}{2}a_1 + \frac{1}{3}a_2 + \frac{1}{4}a_3\,.$$

Inserting the expressions for the coefficients, we obtain

$$\int_0^1 f(t)\,dt \approx \frac{1}{2}(f(0)+f(1)) + \frac{1}{12}(f'(0)-f'(1))\,. \tag{10.1}$$

Let f be defined and differentiable on $[0,n]$, where n is some positive integer. From (10.1) we get the approximation

$$\begin{aligned} \int_{\frac{j-1}{m}}^{\frac{j}{m}} f(t)\,dt &= \int_0^1 f\left(\frac{j-1+t}{m}\right)\frac{dt}{m} \\ &\approx \frac{1}{2m}\left(f\left(\frac{j-1}{m}\right)+f\left(\frac{j}{m}\right)\right) + \frac{1}{12\,m^2}\left(f'\left(\frac{j-1}{m}\right)-f'\left(\frac{j}{m}\right)\right)\,. \end{aligned}$$

Summing over $j = 1, \ldots, mn$, we arrive at the following approximation, known as Woolhouse's formula:

$$\int_0^n f(t)\,dt \approx \frac{1}{m}\sum_{j=1}^{mn-1} f\left(\frac{j}{m}\right) + \frac{f(0)+f(n)}{2\,m} + \frac{f'(0)-f'(n)}{12\,m^2}\,. \tag{10.2}$$

We easily reshape it to obtain the following two relationships:

$$\frac{1}{m}\sum_{j=0}^{mn-1} f\left(\frac{j}{m}\right) \approx \int_0^n f(t)\,dt + \frac{f(0)-f(n)}{2\,m} - \frac{f'(0)-f'(n)}{12\,m^2}\,. \tag{10.3}$$

$$\frac{1}{m}\sum_{j=1}^{mn} f\left(\frac{j}{m}\right) \approx \int_0^n f(t)\,dt - \frac{f(0)-f(n)}{2\,m} - \frac{f'(0)-f'(n)}{12\,m^2}\,. \tag{10.4}$$

Woolhouse's formula is useful in actuarial computations since it provides approximations of discrete time present values by their continuous time analogs, which are easily computed as numerical solutions to differential equations. For instance, for

$$f(t) = {}_tE_x = e^{-\int_0^t (r+\mu_{x+s})\,ds}$$

we have

$$f'(t) = -{}_tE_x\,(r+\mu_{x+t})\,,$$

and so

$$\ddot{a}^{(m)}_{x\,\overline{n}|} \approx \bar{a}_{x\,\overline{n}|} + \frac{1}{2\,m}(1 - {}_nE_x) + \frac{1}{12\,m^2}((r+\mu_x) - {}_nE_x(r+\mu_{x+n}))\,,$$

$$a^{(m)}_{x\,\overline{n}|} \approx \bar{a}_{x\,\overline{n}|} - \frac{1}{2\,m}(1 - {}_nE_x) + \frac{1}{12\,m^2}((r+\mu_x) - {}_nE_x(r+\mu_{x+n}))\,.$$

10.2 More about salary related benefits

Adopting the model assumptions and the notation in the exercises file '305exerc', Paragraph O in the section on Surplus, we will discuss how to evaluate contributions and benefits under various schemes with salary dependent payments. The contributions are usually a certain fraction of the salary. Assume they are payable continuously as an m year life annuity at rate $\pi\,S(t)$ at time $t \in (0, m)$.

A. Expected value of the discounted contributions. Given that the economy is in state e at time 0, the expected present value of the total premiums paid is π times

$$\mathbb{E}\left[\int_0^m e^{\int_0^\tau (a(s)-r(s))\,ds}\,{}_\tau p_x\,d\tau \,\middle|\, Y(0) = e\right]\,.$$

To determine this expected value, we start from the conditional expected present value, given the history $\mathcal{F}_t$ at time $t \in [0, m]$:

$$\mathbb{E}\left[\int_0^m e^{\int_0^\tau (a(s)-r(s))\,ds}\,{}_\tau p_x\,d\tau \,\middle|\, \mathcal{F}_t\right] = \int_0^t e^{\int_0^\tau (a(s)-r(s))\,ds}\,{}_\tau p_x\,d\tau$$
$$+\, e^{\int_0^t (a(s)-r(s))\,ds}\,\mathbb{E}\left[\int_t^m e^{\int_t^\tau (a(s)-r(s))\,ds}\,{}_\tau p_x\,d\tau \,\middle|\, \mathcal{F}_t\right].$$

Due to the Markov property, the expected value on the right hand side depends only on the time t and the current state $Y(t)$. Thus, we need to determine the functions

$$W_e(t) = \mathbb{E}\left[\int_t^m e^{\int_t^\tau (a(s)-r(s))\,ds}\,{}_\tau p_x\,d\tau \,\middle|\, Y(t) = e\right], \tag{10.1}$$

$t \in [0, m]$. Preparing for the backward argument, decompose the expression under the expectation sign into what pertains to the small interval $(t, t + dt]$ and what pertains to the interval $(t + dt, m]$:

$${}_t p_x\,dt + e^{(a(t)-r(t))\,dt} \int_{t+dt}^m e^{\int_{t+dt}^\tau (a(s)-r(s))\,ds}\,{}_\tau p_x\,d\tau + o(dt)\,.$$

Conditioning on what happens in $(t, t + dt]$, gives

$$W_e(t) = (1 - \lambda_{e\cdot} dt)\Big({}_t p_x\,dt + e^{(a_e - r_e)\,dt} W_e(t + dt)\Big) + \sum_{f;f\neq e} \lambda_{ef}\,dt\,W_f(t) + o(dt)\,.$$

We arrive at the differential equations

$$\frac{d}{dt} W_e(t) = W_e(t)(r_e - a_e) - {}_t p_x - \sum_{f;f\neq e} \lambda_{ef}(W_f(t) - W_e(t))\,, \tag{10.2}$$

$t \in (0, m)$, subject to the side conditions

$$W_e(m-) = 0\,, \tag{10.3}$$

$e = 1, \ldots, J^Y$. Solving these equations, we determine the value of $W_e(0)$, which was the original problem.

The expected discounted contributions need to be equated to the expected discounted benefits. We will discuss some cases.

B. Expected value of the discounted benefits; The final salary scheme. The benefits are an m years deferred n years temporary life annuity payable continuously at rate $k\,S(m) = k\,e^{\int_0^m a(s)\,ds}$. Given that the economy is in state e at time 0, the expected present value of the benefits is k times

$$\mathbb{E}\left[e^{\int_0^m a(s)\,ds} \int_m^{m+n} e^{-\int_0^\tau r(s)\,ds}\,{}_\tau p_x\,d\tau \,\middle|\, Y(0) = e\right]. \tag{10.4}$$

Start from

$$\mathbb{E}\left[e^{\int_0^m a(s)\,ds} \int_m^{m+n} e^{-\int_0^\tau r(s)\,ds}{}_\tau p_x \, d\tau \middle| \mathcal{F}_t \right] . \tag{10.5}$$

For $m \le t \le m+n$ this expression can suitably be decomposed as

$$e^{\int_0^m a(s)\,ds} \left(\int_m^t e^{-\int_0^\tau r(s)\,ds}{}_\tau p_x \, d\tau + e^{-\int_0^t r(s)\,ds} \mathbb{E}\left[\int_t^{m+n} e^{-\int_t^\tau r(s)\,ds}{}_\tau p_x \, d\tau \middle| \mathcal{F}_t \right] \right) .$$

We need to determine the functions

$$W_e'(t) = \mathbb{E}\left[\int_t^{m+n} e^{-\int_t^\tau r(s)\,ds}{}_\tau p_x \, d\tau \middle| Y(t) = e \right] , \tag{10.6}$$

$t \in [m, m+n]$. The problem is essentially the same as that of determining the functions W_e in (10.1); the functions W_e' are solution to

$$\frac{d}{dt} W_e'(t) = W_e'(t) r_e - {}_t p_x - \sum_{f; f \neq e} \lambda_{ef} (W_f'(t) - W_e'(t)) , \tag{10.7}$$

$t \in (m, m+n)$, subject to

$$W_e'(m+n-) = 0 . \tag{10.8}$$

For $0 \le t \le m$ we decompose (10.5) as

$$e^{\int_0^t (a(s)-r(s))\,ds} \mathbb{E}\left[e^{\int_t^m (a(s)-r(s))\,ds} \int_m^{m+n} e^{-\int_m^\tau r(s)\,ds}{}_\tau p_x \, d\tau \middle| \mathcal{F}_t \right] .$$

Thus we need to determine the functions

$$W_e''(t) = \mathbb{E}\left[e^{\int_t^m (a(s)-r(s))\,ds} \int_m^{m+n} e^{-\int_m^\tau r(s)\,ds}{}_\tau p_x \, d\tau \middle| Y(t) = e \right] . \tag{10.9}$$

Straightforwardly, they are the solution to

$$\frac{d}{dt} W_e''(t) = W_e''(t)(r_e - a_e) - \sum_{f; f \neq e} \lambda_{ef} (W_f''(t) - W_e''(t)) , \tag{10.10}$$

$t \in (0, m)$, subject to the conditions

$$W_e''(m-) = W_e'(m) , \tag{10.11}$$

where the W_e' were obtained as the solutions to (10.7) – (10.8).

C. Expected value of the discounted benefits; The average salary scheme. The benefits are an m years deferred n years temporary life annuity payable continuously at rate $k \, \frac{1}{m} \int_0^m S(\tau)\, d\tau = k \, \frac{1}{m} \int_0^m e^{\int_0^\tau a(s)\,ds}\, d\tau$. Given that the economy is in state e at time 0, the expected present value of the benefits is k/m times

$$\mathbb{E}\left[\int_0^m e^{\int_0^\tau a(s)\,ds}\,d\tau \int_m^{m+n} e^{-\int_0^\tau r(s)\,ds}\,{}_\tau p_x\,d\tau \,\middle|\, Y(0)=e\right].$$

Start from

$$\mathbb{E}\left[\int_0^m e^{\int_0^\tau a(s)\,ds}\,d\tau \int_m^{m+n} e^{-\int_0^\tau r(s)\,ds}\,{}_\tau p_x\,d\tau \,\middle|\, \mathcal{F}_t\right]. \tag{10.12}$$

For $m \le t \le m+n$ this expression decomposes as

$$\int_0^m e^{\int_0^\tau a(s)\,ds}\,d\tau \left(\int_m^t e^{-\int_0^\tau r(s)\,ds}\,{}_\tau p_x\,d\tau + e^{-\int_0^t r(s)\,ds}\mathbb{E}\left[\int_t^{m+n} e^{-\int_t^\tau r(s)\,ds}\,{}_\tau p_x\,d\tau \,\middle|\, \mathcal{F}_t\right].\right)$$

Thus, the functions we need to determine are precisely the W_e' in (10.6), which are the solution to (10.7) – (10.8).

For $0 \le t \le m$ we decompose (10.12) as

$$\begin{aligned}&\int_0^t e^{\int_0^\tau a(s)\,ds}\,d\tau\, e^{-\int_0^t r(s)\,ds}\mathbb{E}\left[e^{-\int_t^m r(s)\,ds}\int_m^{m+n} e^{-\int_m^\tau r(s)\,ds}\,{}_\tau p_x\,d\tau \,\middle|\, \mathcal{F}_t\right]\\ &+e^{\int_0^t (a(s)-r(s))\,ds}\mathbb{E}\left[\int_t^m e^{\int_t^\tau a(s)\,ds}\,d\tau\, e^{-\int_t^m r(s)\,ds}\int_m^{m+n} e^{-\int_m^\tau r(s)\,ds}\,{}_\tau p_x\,d\tau \,\middle|\, \mathcal{F}_t\right].\end{aligned}$$

Thus, we need to determine the functions

$$\begin{aligned}W_e''(t) &= \mathbb{E}\left[e^{-\int_t^m r(s)\,ds}\int_m^{m+n} e^{-\int_m^\tau r(s)\,ds}\,{}_\tau p_x\,d\tau \,\middle|\, Y(t)=e\right],\\ W_e'''(t) &= \mathbb{E}\left[\int_t^m e^{\int_t^\tau a(s)\,ds}\,d\tau\, e^{-\int_t^m r(s)\,ds}\int_m^{m+n} e^{-\int_m^\tau r(s)\,ds}\,{}_\tau p_x\,d\tau \,\middle|\, Y(t)=e\right],\end{aligned}$$

$t \in [0,m]$. The W_e'' are essentially the same as those in (10.9) and are the solution to

$$\frac{d}{dt}W_e''(t) = W_e''(t)r_e - \sum_{f;f\neq e}\lambda_{ef}(W_f''(t) - W_e''(t)), \tag{10.13}$$

$t \in (0,m)$, subject to the conditions

$$W_e''(m-) = W_e'(m). \tag{10.14}$$

To determine the W_e''', reshape the expression under the expectation as

$$\begin{aligned}&dt\, e^{-\int_t^m r(s)\,ds}\int_m^{m+n} e^{-\int_m^\tau r(s)\,ds}\,{}_\tau p_x\,d\tau\\ &+e^{(a(t)-r(t))\,dt}\int_{t+dt}^m e^{\int_{t+dt}^\tau a(s)\,ds}\,d\tau\, e^{-\int_{t+dt}^m r(s)\,ds}\int_m^{m+n} e^{-\int_m^\tau r(s)\,ds}\,{}_\tau p_x\,d\tau.\end{aligned}$$

Using the backward construction, we find

$$W_e'''(t) = (1-\lambda_e.dt)\left(dt\, W_e''(t) + e^{(a_e-r_e)\,dt}W_e'''(t+dt)\right) + \sum_{f;\, f\neq e}\lambda_{ef}W_f'''(t+dt)\,,$$

and obtain the differential equations

$$\frac{d}{dt}W_e'''(t) = W_e'''(t)\,(r_e - a_e) - W_e''(t) - \sum_{f; f\neq e}\lambda_{ef}(W_f'''(t) - W_e'''(t))\,, \quad (10.15)$$

$t \in (0, m)$, subject to

$$W_e'''(m-) = 0\,. \quad (10.16)$$

The differential equations (10.13) and (10.15) are to be solved simultaneously.

D. Expected value of the discounted benefits; The final salary scheme with guarantee. We modify the situation in Paragraph B so that the benefit is payable at rate k times $S(m) \vee g$, where g is a guaranteed minimum pension rate. We need to determine

$$\mathbb{E}\left[\left(e^{\int_0^m a(s)\,ds} \vee g\right)\int_m^{m+n} e^{-\int_0^\tau r(s)\,ds}\,{}_\tau p_x\,d\tau\,\middle|\, Y(0) = e\right]\,.$$

Start from

$$\mathbb{E}\left[\left(e^{\int_0^m a(s)\,ds} \vee g\right)\int_m^{m+n} e^{-\int_0^\tau r(s)\,ds}\,{}_\tau p_x\,d\tau\,\middle|\, \mathcal{F}_t\right]\,. \quad (10.17)$$

For $m \le t \le m+n$ the problem is the same as in Paragraph B, and we arrive at the W_e' determined by (10.7) – (10.8).

For $0 \le t \le m$, rearrange (10.17) as

$$e^{-\int_0^t r(s)\,ds}\mathbb{E}\left[\left(e^{\int_0^t a(s)\,ds}e^{\int_t^m a(s)\,ds} \vee g\right)e^{-\int_t^m r(s)\,ds}\int_m^{m+n} e^{-\int_m^\tau r(s)\,ds}\,{}_\tau p_x\,d\tau\,\middle|\, \mathcal{F}_t\right]\,.$$

Due to the non-linearity of the "floor" function $x \vee g$, we realize that we can no longer get the function

$$U(t) = e^{\int_0^t a(s)\,ds}$$

outside the expectation. Using the Markov property, by which the past and the future of the driving Markov process are conditionally independent for fixed present state, we realize that the problem is to determine the functions

$$W_e''(t,u) = \mathbb{E}\left[\left(ue^{\int_t^m a(s)\,ds} \vee g\right)e^{-\int_t^m r(s)\,ds}\int_m^{m+n} e^{-\int_m^\tau r(s)\,ds}\,{}_\tau p_x\,d\tau\,\middle|\, Y(t) = e\right]\,.$$

Preparing for the backward argument, decompose the expression under the expectation into what pertains to the small interval $(t, t+dt]$ and what pertains to the interval $(t+dt, m+n]$:

$$\left(ue^{a(t)\,dt}e^{\int_{t+dt}^m a(s)\,ds} \vee g\right)e^{-r(t)\,dt}e^{-\int_{t+dt}^m r(s)\,ds}\int_m^{m+n} e^{-\int_m^\tau r(s)\,ds}\,{}_\tau p_x\,d\tau\,.$$

Condition on what happens in $(t, t+dt]$:

$$\begin{aligned}
W''_e(t,u) &= (1-\lambda_{e\,.}\,dt)e^{-r_e\,dt}W''_e(t+dt, ue^{a_e\,dt}) + \sum_{f;\,f\neq e}\lambda_{ef}\,dt\,W''_f(t,u) + o(dt) \\
&= (1-\lambda_{e\,.}\,dt)(1-r_e\,dt)W''_e(t+dt, u+ua_e dt) + \sum_{f;\,f\neq e}\lambda_{ef}\,dt\,W''_f(t,u) + o(dt) \\
&= (1-\lambda_{e\,.}\,dt - r_e\,dt)\left(W''_e(t,u) + \frac{\partial}{\partial t}W''_e(t,u)\,dt + \frac{\partial}{\partial u}W''_e(t,u)\,u\,a_e\,dt\right) \\
&\quad + \sum_{f;\,f\neq e}\lambda_{ef}\,dt\,W''_f(t,u) + o(dt)\,.
\end{aligned}$$

We arrive at the partial differential equations

$$\frac{\partial}{\partial t}W''_e(t,u) = W''_e(t,u)r_e - \frac{\partial}{\partial u}W''_e(t,u)\,u\,a_e - \sum_{f;f\neq e}\lambda_{ef}(W''_f(t,u) - W''_e(t,u)), \tag{10.18}$$

$t \in (0,m)$, subject to the conditions

$$W''_e(m-,u) = (u \vee g)\,W'_e(m)\,. \tag{10.19}$$

10.3 Relationships between present values of endowments, annuities, and assurances

Consider the standard multi-state policy with payment function B of the form

$$dB(t) = \sum_g I_g(t)\,dB_g(t) + \sum_{g\neq h} b_{gh}(t)\,dN_{gh}(t)\,. \tag{10.20}$$

Suppose the investment portfolio of the insurance company bears interest with intensity $r(t)$ at time t. Using Itǒ's formula we get

$$\begin{aligned}
d\left(e^{-\int_0^\tau r}I_j(\tau)\,B_j(\tau)\right) &= e^{-\int_0^\tau r}(-r(\tau)\,d\tau)I_j(\tau)\,B_j(\tau) \\
&\quad + e^{-\int_0^\tau r}dI_j(\tau)\,B_j(\tau) \\
&\quad + e^{-\int_0^\tau r}I_j(\tau)\,dB_j(\tau) \\
&= -\,e^{-\int_0^\tau r}I_j(\tau)B_j(\tau)\,r(\tau)\,d\tau \\
&\quad + e^{-\int_0^\tau r}(dN_{.\,j}(\tau) - dN_{j\,.}(\tau))B_j(\tau) \\
&\quad + e^{-\int_0^\tau r}I_j(\tau)\,dB_j(\tau)\,.
\end{aligned}$$

Here we have used the identity

$$dI_j(\tau) = dN_{.\,j}(\tau) - dN_{j\,.}(\tau)\,,$$

which states the obvious fact that $I_j(t)$, which assumes only values 0 and 1, makes a jump from 0 to 1 upon a transition into state j and a jump of -1 upon a transition out of state j. We have also used the fact that the three functions involved here cannot jump at the same time with any positive probability: The function $e^{-\int_0^\tau r}$ is continuous and causes no problem. The function $B_j(\tau)$ is deterministic and has at most a finite number of discontinuities at fixed points of time in any finite time interval. The probability that the Markov chain makes a transition (causing a jump in two of the indicator functions) at any of those fixed times is 0.

Now, integrating the differential above from 0 to n, we arrive at

$$\begin{aligned} e^{-\int_0^n r} I_j(n)\, B_j(n) - I_j(0)\, B_j(0) &= -\int_0^n e^{-\int_0^\tau r} I_j(\tau) B_j(\tau)\, r(\tau)\, d\tau \\ &\quad + \sum_{k;\, k\neq j} \int_0^n e^{-\int_0^\tau r} B_j(\tau)\, dN_{kj}(\tau) \\ &\quad - \sum_{k;\, k\neq j} \int_0^n e^{-\int_0^\tau r} B_j(\tau)\, dN_{jk}(\tau) \\ &\quad + \int_0^n e^{-\int_0^\tau r} I_j(\tau)\, dB_j(\tau)\,. \end{aligned}$$

This is a general relationship between present values of endowments, annuities, and assurances. It holds for all payment functions B_j (also path-dependent) and all interest rate processes (also stochastic), and is true with probability 1 (i.e. does not have anything to do with expected values).

Consider the simple model with two states 0 ('alive') and 1 ('dead'), $Z(0) = 0$ (hence $I_0(0) = 1$), and constant interest rate r. Taking $B_0(t) \equiv 1$, and forming expectation, we get the well-known result

$${}_nE_x - 1 = -r\, \bar{a}_{x\,\overline{n|}} - \bar{A}_{\overset{1}{x}\,\overline{n|}}\,.$$

Taking $B_0(t) = 0 \vee (t \wedge n)$, we obtain

$$n\, {}_nE_x = -r\, (\bar{I}\bar{a})_{x\,\overline{n|}} - (\bar{I}\bar{A})_{\overset{1}{x}\,\overline{n|}} + \bar{a}_{x\,\overline{n|}}\,.$$

More generally, with the self-explaining notation

$$E_{tu}^{ij} = e^{-\int_t^u r(s)\, ds} p_{ij}(t,u), \tag{10.21}$$

$$\bar{a}_{tu}^{ij}(dB) = \int_t^u e^{-\int_t^\tau r(s)\, ds} p_{ij}(t,\tau)\, dB(\tau), \tag{10.22}$$

$$\bar{A}_{tu}^{ijk}(b) = \int_t^u e^{-\int_t^\tau r(s)\, ds} p_{ij}(t,\tau)\, \mu_{jk}(\tau)\, b(\tau)\, d\tau, \tag{10.23}$$

forming expectation in the result above gives:

$$\begin{aligned} E_{0n}^{ij}\, B_j(n) - \delta_{ij} B_j(0) &= -\bar{a}_{0n}^{ij}(r(t)\, B_j(t)\, dt) \\ &\quad + \sum_{k;\, k\neq j} \bar{A}_{0n}^{ikj}(B_j) - \sum_{k;\, k\neq j} \bar{A}_{0n}^{ijk}(B_j) \\ &\quad + \bar{a}_{0n}^{ij}(dB_j)\,. \end{aligned}$$

Relationships of this type are not particularly useful for computational purposes.

10.4 Multi-life functions in the Markov chain model

A. The example: A Markov chain model for three lives. Chapter 6 "Multi-life insurances" in the 'lifebook' is about insurance products with payments dependent on the number of survivors in a given body of r lives. From a practical point of view (i.e. in terms of market shares) the joint-life status and last-survivor status for two lives (typically husband and wife) are the most important. The general q survivor status for r lives is mainly of theoretical interest or, rather, used to have theoretical interest until some 35 years ago when all these situations (plus a vast range of other situations) were embraced in the stochastic process model for the general multi-state insurance policy. Figure 10.1 shows the relevant Markov chain for analysis of insurances involving three lives (x), (y), and (z).

B. Payments dependent on the number of survivors. In the standard Markov model one works with the transition probabilities $p_{jk}(t,u)$ and the intensities $\mu_{jk}(t)$, and the basic actuarial functions are those defined in (10.21) – (10.23). Functions defined in Chapter 6 in the 'lifebook' translate to e.g.:

$$
\begin{aligned}
{}_tp_{xyz} &= p_{00}(0,t)\,, \\
{}_tp_{\overline{xyz}} &= 1 - p_{07}(0,t)\,, \\
{}_tp_{\overset{\,2}{\overline{xyz}}} &= p_{00}(0,t) + p_{01}(0,t) + p_{02}(0,t) + p_{03}(0,t)\,, \\
{}_nE_{xyz} &= E^{00}_{0n}\,, \\
\bar{a}_{xyz\,\overline{n}|} &= \bar{a}^{00}_{0n}\,, \\
\bar{A}_{\overset{1}{\overline{xyz}}\,\overline{n}|} &= \bar{A}^{47}_{0n} + \bar{A}^{57}_{0n} + \bar{A}^{67}_{0n}\,, \\
\bar{A}_{\overset{\overset{1}{2}}{\overline{xyz}}\,\overline{n}|} &= \bar{A}^{14}_{0n} + \bar{A}^{15}_{0n} + \bar{A}^{24}_{0n} + \bar{A}^{26}_{0n} + \bar{A}^{35}_{0n} + \bar{A}^{36}_{0n}\,.
\end{aligned}
$$

The last example demonstrates a difficulty with standard actuarial notation. Let is discuss computation of this quantity. The formulas in Paragraph 6.1E do not offer an easy way to compute it. The Markov model sketched in Fig. 10.1 does. Plug the particulars of the situation into the program 'prores1.pas': 8 states numbered from 0 to 7, force of interest and mortality as specified in the valuation basis, term of the contract n years, and benefits being a unit payable upon transitions $1 \to 4$, $1 \to 5$, $2 \to 4$, $2 \to 6$, $3 \to 5$, $3 \to 6$. Run the program, and you are done. The Markov model allows of dependence between the individual remaining life lengths – all it takes is to specify suitable transition intensities (see Section 7.9 in the lifebook). The approach in Chapter 6 is unable to capture such features of the situation.

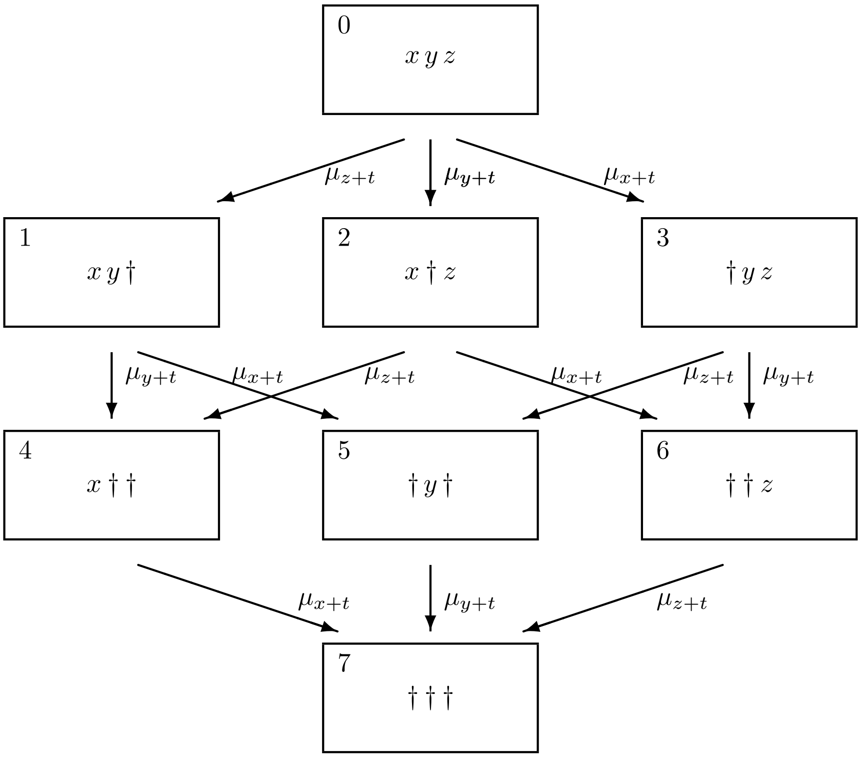


Figure 10.1: Markov model for three lives. A † indicates that the status is dead.

C. Contingent functions: Payments dependent on the order of deaths. Actuarial values of multi-life insurances with payments dependent, not only on the number of the survivors, but also on the order in which the members of the group die, are called contingent functions. Some of those are treated in Unit 1 of the Core Reading notes. They are of course easily accommodated in the Markov chain framework. For instance, consider $\bar{A}_{\overset{1}{(xy)}\,z}$, the actuarial value payable upon death of the joint life (xy) if (z) is still alive: This is a unit payable upon transitions $0 \to 2$ and $0 \to 3$. (Actually, only states 0 - 4 were needed in this case - a multiple decrement model.) The function $\bar{A}_{x\overset{2}{y}\,\overline{n|}}$ represents a unit payable upon the death of (y) if (x) is already dead and $\overset{1}{\overline{n|}}$ is still "alive", that is, if (y) dies after (x) within n years. Referring to Fig. 7.4 in the lifebook, taking husband to be (x) and wife to be (y), this is a unit payable upon transition $1 \to 3$ before time n. Just to illustrate the notation, the function $\bar{A}_{x\overset{2}{y}\,\overline{n|}}$ is the single premium for a unit payable upon the death of (y) if it expires second of the three statuses listed; upon transition $1 \to 3$ before time n and upon transition $0 \to 2$ after time n. Discuss how to accommodate functions like ${}_{m|n}\bar{a}_{\overline{xy}}$, $\bar{a}_{(xy)\,\overline{n}|}$, and $\bar{a}_{x|y}$ in the Markov model. Notice that annuity payments are usually made at discrete times in practice, so that we would work with e.g. $a_{x\,\overline{n}|}$ (annual payments) instead of $\bar{a}_{x\,\overline{n}|}$.

10.5 Alterations to the contract

For a background on this topic, see Unit 8 in the Core reading notes. For the sake of concreteness, let us consider an m-year deferred n-year life annuity (pension) at rate b against premium payable continuously at level rate c in the deferred period. The policy-holder is x years at the time of issue of the contract, and transition intensities are policy duration select as indicated in the figure. Take the interest rate r to be constant (just a matter of notation). We assume that the scheme leaves the insured with the following *surrender paid-up options.* If the policy is *surrendered* (cancelled) at some time $t \in (0, m)$, then the policy-holder receives a *surrender value* $w(t)$. If the policy *lapses* (premiums cease to be paid, but the policy remains in force) at some time $t \in (0, m)$, then the pension benefit will be reduced to a certain fraction $q(t)$ of what is specified in the policy. The altered pension rate, $q(t)\,b$, is called the *paid-up pension rate.* We will discuss how to design the functions $w(t)$ and $q(t)$ and will do so in the framework of the multi-state Markov chain model sketched in Figure 10.2.

In state 1 (surrendered) the policy has expired and, trivially, no reserve is to be provided; $V_1(t) = 0$ for all t.

If the policy is in state 2 (lapsed) at policy duration t and state duration s, then the reserve is

$$V_2(s,t) = q(t-s)\,\bar{V}_2(t)\,, \tag{10.24}$$

where $\bar{V}_2(t)$ is the premium reserve in respect of the full benefit specified in the contract:

$$\begin{aligned}
\bar{V}_2(t) &= b\int_{t\vee m}^{m+n} e^{-\int_t^\tau (r+\nu_{[x]+u})\,du}\,d\tau \\
&= \begin{cases} b\,e^{-\int_t^m (r+\nu_{[x]+u})\,du}\int_m^{m+n} e^{-\int_m^\tau (r+\nu_{[x]+u})\,du}\,d\tau\,, & 0 \le t < m\,, \\ b\int_t^{m+n} e^{-\int_t^\tau (r+\nu_{[x]+u})\,du}\,d\tau\,, & m \le t \le m+n\,. \end{cases}
\end{aligned}$$

The "baseline" reserve $\bar{V}_2(t)$ is best computed as the solution to the simple differential equation

$$\frac{d}{dt}\bar{V}_2(t) \;=\; \bar{V}_2(t)(r+\nu_{[x]+t}) \;-\; b\,1_{(m,m+n)}(t)\,, \tag{10.25}$$

subject to the condition

$$\bar{V}_2(m+n) \;=\; 0\,.$$

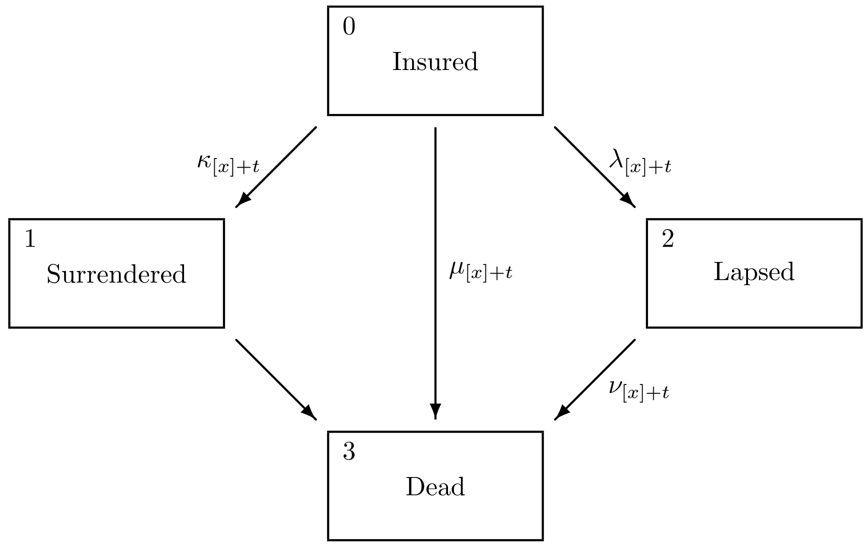


Figure 10.2: A model for a single-life insurance with surrender and lapse options.

With these specifications of the surrender and paid up benefits, the reserve in state 0 is given by the differential equation

$$\frac{d}{dt}V_0(t) = \begin{cases} V_0(t)\,r + c - \mu_{[x]+t}(-V_0(t)) - \kappa_{[x]+t}(w(t) - V_0(t)) \\ \qquad -\lambda_{[x]+t}(q(t)\bar{V}_2(t) - V_0(t))\,,\ 0 < t < m\,, \\ V_0(t)\,r - b - \mu_{[x]+t}(-V_0(t))\,,\ m < t < m+n\,. \end{cases} \tag{10.26}$$

At any time and in any state the current reserve is the insurer's expected future liabilities under the contract. In more suggestive terms, it is the insurer's debt to the insured. Therefore, upon surrender the reserve should be paid back to the insured cash:

$$w(t) = V_0(t)\,. \tag{10.27}$$

Similarly, upon lapsing the current reserve should be seen as a single premium for an equivalent fraction $q(t)$ of the contractual benefits. By the principle of equivalence, $q(t)$ is given by $V_0(t) = q(t)\,\bar{V}_2(t)$, or

$$q(t) = \frac{V_0(t)}{\bar{V}_2(t)}\,. \tag{10.28}$$

With $w(t)$ and $q(t)$ given by (10.27) and (10.28), the relation (10.26) reduces to

$$\frac{d}{dt}V_0(t) = V_0(t)\,(r + \mu_{[x]+t}) + c$$

for $t \in (0, m)$, which is nothing but the differential equation for the reserve in state 0 in the situation without surrenders and lapses. This is quite natural since the surrendered and lapsed policies take with them just what the scheme owes them - nothing is bequeathed to those who stay in the scheme. In particular, the premium rate c determined by the equivalence requirement $V_0(0) = 0$, is not affected by the surrender and paid-up options.

10.6 Prospective and retrospective reserves

A. Definition. In a quite context-free set-up, let B be the payment function generated by benefits less premiums on an insurance policy issued at time 0 and terminating at time n. Payments are currently invested (premiums deposited and benefits withdrawn) in an account that bears interest at rate $r(t)$ at time t. The balance of the company's account at time t is the sum of all past and present payments accumulated with interest,

$$U(t) = \int_{0-}^{t} e^{\int_\tau^t r(s)\,ds} d(-B)(\tau)\,,$$

where $\int_{0-}^{t} = \int_{[0,\,t]}$. The company's discounted (strictly) future net liability at time t is

$$V(t) = \int_{t}^{n} e^{-\int_t^\tau r(s)\,ds} dB(\tau)\,. \tag{10.29}$$

Let $\mathcal{H}_t$ represent the complete history of the policy up to and including time t and introduce the corresponding "flow of information" $\mathbf{H} = \{\mathcal{H}_t\}_{t\geq 0}$. In mathematical terms, $\mathcal{H}_t$ is the sigma-algebra of policy events that are observable by time t, and $\mathbf{H}$ is a filtration (meaning $\mathcal{H}_s \subset \mathcal{H}_t$ if $s < t$ – no information is thrown away).

At any time t the current balance $U(t)$ is known, that is, fully determined by the information $\mathcal{H}_t$. Mathematically speaking, the process U is adapted to the filtration $\mathbf{H}$. The future liability $V(t)$ is not known at time t in general. The reserve that has to be provided to meet this liability can, of course, only be based on the available information $\mathcal{H}_t$. Legislation lays down that the reserve should be the mean value of $V(t)$ based on what is currently known. We call it the *prospective reserve* (based on full information), denote it by $V_{\mathbf{H}}(t)$, and define it precisely as

$$V_{\mathbf{H}}(t) = \mathbb{E}\left[\,V(t)\,|\,\mathcal{H}_t\right]\,. \tag{10.30}$$

The cash balance, which is the money that in principle should cover the future liability, can suitably be called the *retrospective reserve* (based on full information). For the sake of symmetry, we might introduce $U_{\mathbf{H}}(t) = \mathbb{E}\left[\,U(t)\,|\,\mathcal{H}_t\right]$, which is just $U(t)$ itself.

The qualifying term "based on full information" suggests that we may create different notions of reserves based on more summary information. Thus, let $\mathcal{H}'_t$

be some partial description of the history of the policy by time t. Mathematically speaking, it is a sub-sigma-algebra of $\mathcal{H}_t$; $\mathcal{H}'_t \subset \mathcal{H}_t$. The flow of information provided by these excerpts of the full policy history is denoted by $\mathbf{H}' = \{\mathcal{H}'_t\}_{t\geq 0}$. It need not be a filtration. Analagous to the reserves based on full information, we can now define the reserves based on $\mathbf{H}'$; the prospective $\mathbf{H}'$-reserve

$$V_{\mathbf{H}'}(t) = \mathbb{E}\left[V(t) \,|\, \mathcal{H}'_t \right] , \tag{10.31}$$

and the retrospective $\mathbf{H}'$-reserve

$$U_{\mathbf{H}'}(t) = \mathbb{E}\left[U(t) \,|\, \mathcal{H}'_t \right] . \tag{10.32}$$

There are various conceivable reasons why reserves may be based on summary information: The policy records may be incomplete because some pieces of information are discarded (by accident or routine); Or it may be decided that the reserves, seen as shares of the total fund, should be averaged out over certain groups of policy-holders according to some sort of solidarity principle.

We have usually been working with what we called the state-wise prospective reserves,

$$V_j(t) = \mathbb{E}\left[V(t) \,|\, Z(t) = j \right] ,$$

$j = 1, \ldots, J$. These are the values of the the prospective reserve $V_{Z(t)}(t)$ based on the information about the current state only, that is, $\mathbf{H}' = \sigma\{Z(t)\}$ (the sigma-algebra generated by $Z(t)$). If the process Z is Markov and the payments B do not depend on the past history of the process, then $V_{Z(t)}(t) = V_{\mathbf{H}}(t)$, of course. This was the justification for concentrating on the state-wise reserves. If the process Z or the payments B should not be "memoryless", then the use of state-wise reserves would represent a sacrifice of information.

The retrospective and prospective $\mathbf{H}'$-reserves are not the same in general. This is quite natural since the very purpose of insurance is to redistribute the total savings in the scheme in a manner that mitigates the risk carried by the individual policyholder. If the principle of equivalence is enforced, then the rule of iterated expectation gives

$$\begin{aligned} 0 &= \mathbb{E}\left[\int_{0-}^{n} e^{-\int_0^\tau r(s)\,ds} dB(\tau) \right] = \mathbb{E}\,\mathbb{E}\left[\int_{0-}^{n} e^{-\int_0^\tau r(s)\,ds} dB(\tau) \,\middle|\, \mathcal{H}'_t \right] \\ &= e^{-\int_t^\tau r(s)\,ds} \mathbb{E}\left[V_{\mathbf{H}'}(t) - U_{\mathbf{H}'}(t) \right] , \end{aligned}$$

hence

$$\mathbb{E}\, U_{\mathbf{H}'}(t) = \mathbb{E}\, V_{\mathbf{H}'}(t) \,. \tag{10.33}$$

Thus, on the average in the (infinitely large) portfolio, the retrospective reserve adequately covers the prospective reserve.

B. An example: insurance of a single life. At time 0 a life (x) (aged x) buys a life insurance policy specifying that the sum b is payable immediately upon death before time n and that premiums are to be contributed continuously with level intensity c throughout the insurance period. As usual we denote the remaining life length of (x) by T_x and the survival function by ${}_tp_x = e^{-\int_0^t \mu_{x+s}\,ds}$, where μ is the force of mortality. Assume that interest rate r is constant. Introduce the indicator of survival to time t, $I(t) = 1[T_x > t]$, and the indicator of death by time time t, $N(t) = 1[T_x > t]$. (The counting process $N(t) = 1-I(t)$ counts "the number of deaths" of the insured.) The possible states of the policy are $0 =$ 'alive' and $1 =$ 'dead', and the state of the policy at time t is just $Z(t) = N(t)$.

Suppose the complete history $\mathbf{H}$ of the policy is continually recorded so that it is known at any time if (x) is alive or dead and, in the latter case, when he died. The retrospective $\mathbf{H}$-reserve at time $t \in [0, n]$ is

$$\begin{aligned} U_{\mathbf{H}}(t) &= \int_0^t e^{(t-\tau)\,r}(c\,I(\tau)\,d\tau \;-\; b\,dN(\tau)) \\ &= \begin{cases} c\int_0^{T_x} e^{(t-\tau)\,r}d\tau \;-\; b\,e^{(t-T_x)\,r}\,, & \text{if } T_x \le t\,, \\ c\int_0^{t} e^{(t-\tau)\,r}d\tau\,, & \text{if } T_x > t\,, \end{cases} \\ &= \begin{cases} c\,\frac{e^{t\,r}-e^{(t-T_x)\,r}}{r} \;-\; b\,e^{(t-T_x)\,r}\,, & \text{if } T_x \le t\,, \\ c\,\frac{e^{t\,r}-1}{r}\,, & \text{if } T_x > t\,. \end{cases} \end{aligned}$$

In the present situation the prospective $\mathbf{H}$-reserve is the same as the state-wise reserve, $V_{\mathbf{H}}(t) = V_{Z(t)}$ (the future is trivial in state 1, and the past is trivial in state 0). The state-wise prospective reserves in the two states are

$$V_0(t) = \int_t^n e^{-(\tau-t)\,r}\;{}_{\tau-t}p_{x+t}\,\{b\mu_{x+\tau} - c\}\,d\tau\,, \tag{10.34}$$

$$V_1(t) = 0\,. \tag{10.35}$$

The state-wise retrospective reserves are

$$U_0(t) = c\int_0^t e^{(t-\tau)\,r}\,d\tau = c\,\frac{e^{t\,r}-1}{r} \tag{10.36}$$

(trivial) and

$$U_1(t) = \frac{1}{1-\,{}_tp_x}\int_0^t e^{(t-\tau)\,r}\{c\,({}_\tau p_x - \,{}_tp_x) - b\,{}_\tau p_x\,\mu_{x+\tau}\}\,d\tau\,. \tag{10.37}$$

The latter is obtained as follows: Conditional on death within time t, the probability of survival to τ is $({}_\tau p_x - \,{}_tp_x)/(1-\,{}_tp_x)$ and the probability of death in $(\tau, \tau + d\tau)$ is ${}_\tau p_x\mu_{x+\tau}d\tau/(1-\,{}_tp_x)$, $0 < \tau < t$.

Assume that the premium is set in accordance with the principle of equivalence. In the present simple situation (10.33) reduces to (recall that $V_1(t) = 0$)

$${}_tp_x\,U_0(t) \;+\; (1-\,{}_tp_x)\,U_1(t) \;=\; {}_tp_x\,V_0(t)\,.$$

The equivalence condition, from which (10.33) was derived, is

$$0 = \int_0^n e^{-\tau\,r}\;{}_\tau p_x\,\{b\mu_{x+\tau} - c\}\,d\tau\,.$$

Upon splitting the integral into $\int_0^n = \int_0^t + \int_t^n$, we find

$$\begin{aligned} V_0(t) &= \frac{1}{e^{-t\,r}\;{}_tp_x}\int_0^t e^{-\tau\,r}\;{}_\tau p_x\,\{c - b\mu_{x+\tau}\}\,d\tau \qquad (10.38)\\ &= \int_0^t e^{\int_\tau^t (r\,+\,\mu_{x+s})\,ds}\,\{c - b\mu_{x+\tau}\}\,d\tau\,. \end{aligned}$$

This is what is called the "retrospective reserve" in the Core Reading notes to Subject 105. It is, however, rather a retrospective formula for the prospective reserve that follows from the principle of equivalence. The notion or retrospective reserve given by (10.38) does not generalize in any obvious way to the multi-state policy.

C. A side-remark: Integral expressions and differential equations for state-wise retrospective reserves. We remind of exercise 18 and consider the standard multi-state insurance policy starting in state 0 at $t = 0$ and with payments of the standard form

$$dB(t) = \sum_g I_g(t)\,b_g(t)\,dt + \sum_{g\neq h} b_{gh}(t)\,dN_{gh}(t)\,,$$

where the state-wise annuity rates b_g and the sums assured b_{gh} are deterministic (independent of the past history of the state process Z). We disregard lump sum annuity payments in $(0, n)$ just to ease notation. The state-wise retrospective reserve in state j at time t is

$$\begin{aligned} U_j(t) &= -\frac{1}{p_{0j}(0,t)}\int_{0-}^t e^{\int_\tau^t r(s)\,ds}\sum_g p_{0g}(0,\tau)\,\{b_g(\tau)\,p_{gj}(\tau,t)\\ &\qquad + \sum_{h;\,h\neq g} b_{gh}(\tau)\,\mu_{gh}(\tau)\,p_{hj}(\tau,t)\}\,d\tau\,. \qquad (10.39) \end{aligned}$$

The explanation is the same as in the special case (10.37) above: given that the process is in state 0 at time 0 and in state j at time t, the conditional probability that the process is in state g in the time interval $(\tau, \tau + d\tau)$ is

$$\frac{p_{0g}(0,\tau)\,p_{gj}(\tau,t)}{p_{0j}(0,t)}\,, \qquad (10.40)$$

and the conditional probability that the process jumps from state g to state h time interval $(\tau, \tau + d\tau)$ is

$$\frac{p_{0g}(0,\tau)\,\mu_{gh}(\tau)\,d\tau\,p_{hj}(\tau,t)}{p_{0j}(0,t)}\,.$$

We obtain differential equations for the functions $U_j(t)$ by differentiating the expression in (10.39). It is easier to work with the functions

$$\tilde{U}_j = e^{-\int_0^t r(s)\,ds}\,p_{0j}(0,t)\,U_j(t)\,,$$

since then t will appear only in two places on the right hand side:

$$\begin{aligned}\tilde{U}_j(t) &= -\int_{0-}^{t} e^{-\int_0^\tau r(s)\,ds} \sum_g p_{0g}(0,\tau)\,\{b_g(\tau)\,p_{gj}(\tau,t) \\ &\quad + \sum_{h;\,h\neq g} b_{gh}(\tau)\,\mu_{gh}(\tau)\,p_{hj}(\tau,t)\}\,d\tau\,.\end{aligned}$$

$$\begin{aligned}\tilde{U}_j(t+dt) &= -\int_{0-}^{t} e^{-\int_0^\tau r(s)\,ds} \sum_g p_{0g}(0,\tau)\,\{b_g(\tau)\,p_{gj}(\tau,t+dt) \\ &\quad + \sum_{h;\,h\neq g} b_{gh}(\tau)\,\mu_{gh}(\tau)\,p_{hj}(\tau,t+dt)\}\,d\tau \\ &\quad - e^{-\int_0^t r(s)\,ds} \sum_g p_{0g}(0,t)\,\{b_g(t)\,p_{gj}(t,t) \\ &\quad + \sum_{h;\,h\neq g} b_{gh}(t)\,\mu_{gh}(t)\,p_{hj}(t,t)\}\,dt\; + o(dt)\,.\end{aligned}$$

Inserting

$$p_{gj}(\tau,t+dt) = p_{gj}(\tau,t)(1-\mu_{j\,\cdot}(t)\,dt)\; + \sum_{i;\,i\neq j} p_{gi}(\tau,t)\,\mu_{ij}(t)\,dt$$

and $p_{gj}(t,t) = \delta_{gj}$, we obtain after some easy algebra the following differential equations:

$$\begin{aligned}\frac{d}{dt}\tilde{U}_j(t) &= \sum_{g;g\neq j} \tilde{U}_g(t)\mu_{gj}(t) \;-\; \tilde{U}_j(t)\,\mu_{j\,\cdot}(t) \\ &\quad - e^{-\int_0^\tau r(s)\,ds}\{p_{0j}(0,t)\,b_j(t) + \sum_{g;\,g\neq j} p_{0g}(0,t)\,b_{gj}(t)\,\mu_{gj}(t)\}\end{aligned} \tag{10.41}$$

They are forward equations starting from the initial conditions

$$\tilde{U}_j(0) \;=\; -\delta_{0j}B_0(0)\,, \tag{10.42}$$

and they are to be solved in parallel with the Kolmogorov forward differential equations for the transition probabilities $p_{0j}(0,t)$ and for the discount function $v(t) = e^{-\int_0^t r(s)\,ds}$ ($\frac{d}{dt}v(t) = -v(t)\,r(t)$). In fact, (10.41) are generalizations of the Kolmogorov forward differential equations for the transition probabilities: in the case with no interest, the retrospective reserve in state j at time t for an endowment of -1 in state i at time $s < t$ is just the conditional probability $\frac{p_{0i}(0,s)\,p_{ij}(s,t)}{p_{0j}(0,t)}$, recall (10.40). Thus $\tilde{U}_j(t)$ is $p_{ij}(s,t)$ times $p_{0i}(0,s)$ (which is independent of t).

The equivalence condition (10.33) can be cast in terms of the retrospective reserves as

$$\sum_{j\in\mathcal{Z}} p_{0j}(0,n)U_j(n) = 0. \tag{10.43}$$

The present theory of reserves is taken from Norberg, R. (1991): Reserves in life and pension insurance. *Scandinavian Actuarial Journal* **1991**, 1-22.

10.7 Population dynamics and pension schemes

Up to this point focus has mainly been on the development of a single insurance policy. Only on a couple of occasions have we been referring to the population of (more or less) similar policies from which our single policy was selected. Firstly, in the context of actuarial evaluation of premiums and reserves we formed expected values perceived as averages over an increasingly large portfolio. Secondly, in the context of statistical inference we worked with data from a portfolio and we stated certain properties about ML estimators that are valid asymptotically as the size of the portfolio increases. In either case the "increasing portfolio" was a purely theoretical notion. We will now resume discussions of the workings of insurance schemes at the macro level of the entire portfolio and, like we did at the micro level of the individual policy, attempt to describe the development over time of the total cash flow generated by all policies under some insurance scheme. Obviously, we will then need to extend the now familiar individual life history model with some description of the mechanisms that bring new participants to the scheme. To this end we will invoke some elements of dynamical population theory and synthesize them with elements of actuarial analysis that are already in place.

Basic definitions. Denoting age by x and calendar time by t, assume that the dynamics of a *closed population* (no migration) are driven by time- and age-specific rates of mortality and fertility denoted by $\mu_t(x)$ and $\phi_t(x)$, respectively. Thus, for a person who is alive and aged x at time t, $\mu_t(x)$ and $\phi_t(x)$ are the mean number of deaths and childbirths, respectively, per time unit. (We might refine the picture by distinguishing between males and females, but that would complicate notation without - for our purposes - adding anything essential to the model. In fact, the present "uni-sex" model is consistent with, and can be obtained from, the "bi-sex" model in an obvious manner.) Let $L_t(x)$ denote the number of citizens at age x or less at time t, and introduce the short-hand $L_t(dx) = L_t(x + dx) - L_t(x)$ for the number of citizens aged between x and $x + dx$ at time t. We assume that $\ell_t(x) = \frac{\partial}{\partial x} L_t(x)$ exists and is continuous for all x and t, so that $L_t(dx) = \ell_t(x)\, dx + o(dx)$. Then $L_t(dx)$ is also the number of citizens aged between $x - dx$ and x and, more generally, in any age interval of length dx around age x. Henceforth we will drop the negligible term $o(dx)$.

Dynamics of the population. The $L_t(dx)$ citizens aged between $x - dx$ and x at time t must be the survivors among those who were born in the time interval $[t - x, t - x + dx)$, see Fig. 10.3.

At time $t-x$ there were $L_{t-x}(dy)$ citizens aged between y and $y+dy$, and they gave birth to $L_{t-x}(dy)\, \phi_{t-x}(y)\, dx$ children in the time interval $[t-x, t-x+dx)$. Summing over all ages y, we find that the total number of child-births in that

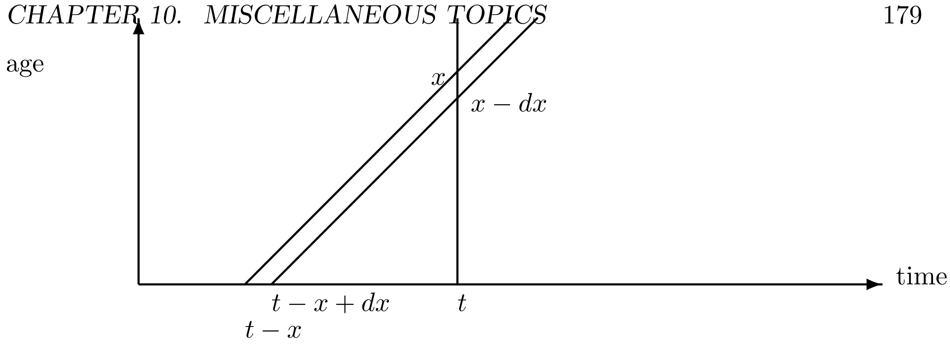


Figure 10.3: Population dynamics

time interval was $\int_0^\infty L_{t-x}(dy)\,\phi_{t-x}(y)\,dx$. These new-born have been subject to the age- and time-dependent mortality rate which, at time $(t-x)+s$ when the survivors have reached age s, is $\mu_{t-x+s}(s)$. Thus, the proportion of survivors to age x is the survival probability $e^{-\int_0^x \mu_{t-x+s}(s)\,ds}$, and we arrive at

$$L_t(dx) = \int_0^\infty L_{t-x}(dy)\,\phi_{t-x}(y)\,dx\,e^{-\int_0^x \mu_{t-x+s}(s)\,ds}\,.$$

Some ideas about pension. Suppose a compulsory pension scheme is introduced at time 0 and that the terms and conditions specify that, at any time $t > 0$, every person aged $x \in [0, z_t)$ is to contribute premium at rate $c_t(x)$ per time unit and every person aged $x \in [z_t, \infty)$ is to receive pension at rate $b_t(x)$ per time unit. Thus, z_t is the retirement age at time t.

Referring to Chapter 2 in BLIM ('Basic Life Insurance Mathematics'), the payment stream A generated by contributions less benefits is given by

$$dA_t = \int_0^{z_t} c_t(x)L_t(dx)\,dt - \int_{z_t}^\infty b_t(x)L_t(dx)\,dt\,.$$

The payments under the scheme are currently deposited on/withdrawn from an account that bears interest at rate r_t per time unit at any time t. Thus, the fund at time t is

$$U_t = \int_0^t e^{\int_\tau^t r_s\,ds} dA_\tau\,,$$

with dynamics

$$\begin{aligned} dU_t &= U_t\,r_t\,dt + dA_t \\ &= U_t\,r_t\,dt + \int_{x=0}^{z_t} c_t(x)L_t(dx)\,dt - \int_{z_t}^\infty b_t(x)L_t(dx)\,dt\,, \end{aligned}$$

starting from $U_0 = 0$. The last two terms on the right could be detailed as

$$\begin{aligned} &\int_{x=0}^{z_t} c_t(x)\int_{y=0}^\infty L_{t-x}(dy)\,\varphi_{t-x}(y)\,dx\,e^{-\int_0^x \mu_{t-x+s}(s)\,ds}\,dt \\ &-\int_{x=z_t}^\infty b_t(x)\int_{y=0}^\infty L_{t-x}(dy)\,\varphi_{t-x}(y)\,dx\,e^{-\int_0^x \mu_{t-x+s}(s)\,ds}\,dt\,. \end{aligned}$$

Pension schemes exist in many forms. Those offered on an individual and voluntary basis by the life insurance offices must be based on the "principle of (individual) equivalence". This means that, for given mortality and interest rates, expected discounted contributions must cover expected discounted benefits so that balance is obtained in a large portfolio. The reason for this is that enrollment in the scheme is voluntary and cannot be anticipated; if the company would seek to cover a deficit by charging premiums in excess of the equivalence rate, then potential new customers would be deterred and existing customers would cancel their policies. The problem with such schemes, with contributions and benefits set out in the contract at the outset, is that one must be able to predict the future development of interest and mortality. During the course we have seen how this problem is tackled in with-profit schemes and in unit-linked schemes.

For compulsory (e.g. national insurance) schemes the situation is quite different. Surpluses and losses may be transferred across groups of participants and also across generations, and it is up to political and governmental bodies to decide when and how in view of experience from the past and predictions about the future. The main characteristics of a given scheme are the functions $c_t(x)$, $b_t(x)$, and z_t and, in particular, the extent to which they can be controlled and adapted to the development of the uncontrollable processes r_t, $\mu_t(x)$, and $\phi_t(x)$. We list some arch-type schemes:

Pay-as-you-go: The functions $c_t(x)$, $b_t(x)$, and z_t are currently chosen such that contributions match benefits at any time. Thus, for all t, $dA_t = 0$ or

$$\int_{x=0}^{z_t} c_t(x) L_t(dx) = \int_{z_t}^{\infty} b_t(x) L_t(dx)\,.$$

Thus, $U_t = 0$ for all t so there is no savings element in the scheme. From a mere solvency point of view, there is no need to predict r, μ, and ϕ or even to know what they are today.

Defined benefits: The functions $c_t(x)$, $b_t(x)$, and z_t are fixed for a certain time period. Contributions must be set sufficiently high to ensure $U_t \geq 0$ for all likely developments of r, μ, and ϕ over the period.

Defined contributions: $c_t(x)$ is fixed for a certain time period. Benefits $b_t(x)$ may be regulated currently to ensure $U_t \geq 0$ over the period.

Some obvious facts can be read out of the formulas. For instance, with time-independent contributions and benefits functions $c(x)$, $b(x)$, and z, we see that a sufficiently big permanent drop in the fertility rates will in due course lead to negative dA_t, and the same goes for a sufficiently big drop in the mortality rates at ages $x > z$.

Many interesting mathematical problems arise from the population dynamics model. We are not going to indulge in such discussions, but after the exam you may try and prove the following statements:

Assume that $\mu_t(x) = \mu(x)$ and $\phi_t(x) = \phi(x)$ are independent of t. Then the size of the population is stationary (which means that $L_t(\infty)$ is independent of t) only if $\int_0^\infty e^{-\int_0^x \mu(s)\,ds}\phi(x)\,dx = 1$ (i.e. one birth per citizen in a lifetime). If $\int_0^\infty e^{-\int_0^x \mu(s)\, ds}\phi(x)\, dx > (<)1$, then the population will grow (decay) exponentially.

10.8 Semi-Markov model and path-dependent payments

A. Model assumptions. So far we have been dealing mainly with situations where payments and transition intensities depend only on the time elapsed since the policy was issued, the so-called *policy duration*, and on the current state. We now extend the analysis to situations where payments and intensities may depend also on on the time elapsed since the last entry into the current state, the so-called *state duration*. Thus, for an n year insurance policy with state space $\{1, 2, \ldots, J\}$, we introduce the following basic entities, where $t \in [0, n]$ denotes policy duration as usual and $s \in [0, t]$ denotes state duration:

$\mu_{jk}(s, t)$, the intensity of transition from state j to state k;

$b_j(s, t)$, the rate at which continuous annuity payments (benefits less premiums); are payable per time unit in state j;

$b_{jk}(s, t)$, the sum assured payable immediately upon transition from state j to state k.

Furthermore, we allow of terminal endowments:

$\Delta B_j(s)$, the terminal endowment in state j at time n.

Lump sum endowments of annuity type before time n can easily be incorporated, but we disregard them here to simplify notation. Assume that the interest rate at time t is deterministic and denote it by $r(t)$. Denote by

$V_j(s, t)$ the reserve in state j at policy duration t and state duration s.

B. Partial differential equations. The state-wise reserves can be constructed straightforwardly by the direct backward argument: Conditioning on whether or not there is a transition out of the current state j in $(t, t + dt]$ and, in case there is, also on the time and the direction of the transition, we get

$$\begin{aligned} V_j(s,t) &= (1 - \mu_{j\,\cdot}(s,t)\, dt) \Big(b_j(s,t)\, dt + e^{-r(t)\, dt} V_j(s + dt, t + dt) \Big) \\ &\quad + \sum_{k;\, k \neq j} \mu_{jk}(s,t)\, dt\ (b_{jk}(s,t) + V_k(0,t)) + o(dt)\,. \end{aligned}$$

Short explanation: Firstly, the probability of no transition in $(t, t + dt]$ is $1 - \mu_{j\,\cdot}(s, t)\, dt$. Given that this happens, there will be annuity payments of $b_j(s, t)\, dt$

in $(t, t + dt]$ and the expected discounted value of payments in $(t + dt, n]$ is $e^{-r(t)\,dt} V_j(s + dt, t + dt)$. Secondly, the probability of a $j \to k$ transition in $(t, t + dt]$ is $\mu_{jk}(s, t)\, dt$. Given this, there will be a payment of $b_{jk}(s, t)$ in $(t, t+dt]$ and the expected discounted value of payments in $(t+dt, n]$ is $V_k(0, t)$. The approximations made here are of order dt^2 or less and are collected in the term $o(dt)$. Now, using the Taylor approximations $e^{-r(t)\,dt} = 1 - r(t)\,dt + o(dt)$ and $V_j(s+dt, t+dt) = V_j(s, t)+\frac{\partial}{\partial s}V_j(s, t)\,dt+\frac{\partial}{\partial t}V_j(s, t)\,dt+o(dt)$ and multiplying out we proceed to

$$\begin{aligned} V_j(s,t) &= b_j(s,t) + (1 - \mu_{j\,\cdot}(s,t)\,dt - r(t)\,dt)V_j(s,t) + \frac{\partial}{\partial s}V_j(s,t)\,dt \\ &\quad + \frac{\partial}{\partial t}V_j(s,t)\,dt + \sum_{k;\,k\neq j} \mu_{jk}(s,t)\,dt\,(b_{jk}(s,t) + V_k(0,t)) + o(dt)\,. \end{aligned}$$

Cancelling $V_j(s, t)$ on both sides of the equation, then dividing by dt and letting $dt \downarrow 0$, we end up with the following first order partial differential equations (PDEs):

$$\begin{aligned} \frac{\partial}{\partial t}V_j(s,t) &= r(t)V_j(s,t) - \frac{\partial}{\partial s}V_j(s,t) - b_j(s,t) \\ &\quad - \sum_{k;k\neq j} \mu_{jk}(s,t)\,(b_{jk}(s,t) + V_k(0,t) - V_j(s,t))\ , \end{aligned} \tag{10.44}$$

$j = 1, \ldots, J$. These are to be solved subject to the conditions

$$V_j(s, n-) = \Delta B_j(s)\,. \tag{10.45}$$

C. Numerical solution to the PDEs Let us first discuss the clear-cut situation where $v(s, t)$ is a vector-valued function defined on [0,S] x [0,T]. To save notation we denote first order partial derivatives w.r.t. s and t by v_s and v_t, respectively, and the second order partial derivatives by v_{ss} , v_{st}, v_{tt}. Assume that v satisfies the linear first order PDE

$$v_t = a + Bv_s + Cv. \tag{10.46}$$

with the condition

$$v(s, T) = w(s). \tag{10.47}$$

Differentiate (10.46) w.r.t. t:

$$v_{tt} = a_t + B_t v_s + Bv_{st} + C_t v + Cv_t. \tag{10.48}$$

Differentiate (10.46) w.r.t. s:

$$v_{st} = a_s + B_s v_s + Bv_{ss} + C_s v + Cv_s. \tag{10.49}$$

The relation (10.46) expresses the first derivative of v in the t direction by v itself and its first derivative in the s direction. Upon substituting (10.46) and (10.49) on the right of (10.48) and rearranging a bit, we get

$$v_{tt} = a_t + Ba_s + Ca + (BC_s + C_t + C^2)v + (B_t + BB_s + BC + CB)v_s + B^2 v_{ss}\,, \tag{10.50}$$

which expresses the second derivative of v in the t direction by v itself and its first two derivatives in the s direction.

Now, Taylor expansion gives

$$v(s, t - \Delta) = v - v_t\Delta + 0.5 v_{tt}\Delta^2 + o(\Delta^3)\,. \tag{10.51}$$

Due (10.46) and (10.50) the expression on the right of (10.51) is a linear expression in v_s and v_{ss}.

Suppose that, for a fixed t, the function $v(\,\cdot\,, t)$ is known. Then we can compute the function $v(\,\cdot\,, t-\Delta)$ as follows. Taylor expansion of second order gives

$$v(s+\Delta, t) = v(s,t) + v_s(s,t)\Delta + \frac{1}{2}v_{ss}(s,t)\Delta^2 + O(\Delta^3)$$

and

$$v(s-\Delta, t) = v(s,t) - v_s(s,t)\Delta + \frac{1}{2}v_{ss}(s,t)\Delta^2 + O(\Delta^3)\,.$$

Upon forming the difference between these two equations (and rearranging a bit), we obtain

$$v_s(s,t) = \frac{v(s+\Delta,t) - v(s-\Delta,t)}{2\Delta} + O(\Delta^2)\,.$$

Upon forming the sum of the two equations, we obtain

$$v_{ss}(s,t) = \frac{v(s+\Delta,t) - 2v(s,t) + v(s-\Delta,t)}{\Delta^2} + O(\Delta)\,.$$

Thus, we can approximate $v_s(s,t)$ and $v_{ss}(s,t)$ by the finite differences

$$v_s^*(s,t) = \frac{v(s+\Delta,t) - v(s-\Delta,t)}{2\Delta}\,, \tag{10.52}$$

$$v_{ss}^*(s,t) = \frac{v(s+\Delta,t) - 2v(s,t) + v(s-\Delta,t)}{\Delta^2}\,, \tag{10.53}$$

with errors of orders $O(\Delta^2)$ and $O(\Delta)$, respectively. Inserting these approximations in (10.46) and (10.50), produces errors of orders $O(\Delta^2)$ and $O(\Delta)$, respectively, and thus an error of order $O(\Delta^3)$ in (10.51).

Use this method backwards on a grid $\{0, \Delta, 2\Delta, ..., S\} \times \{0, \Delta, 2\Delta, ..., T\}$ with step-length Δ, starting from the known condition (10.47). It can be shown that the global error is of order $O(\Delta^2)$.

In our situation the function is defined on the triangle $\{(s,t);\, 0 \le s \le t \le n\}$ rather than on a rectangle $\{(s,t);\, 0 \le s \le S,\, 0 \le t \le T\}$. Moreover the PDEs are what is called *shifted*, which means that it involves values of the unknown functions outside the current state s (the $V_k(0,t)$ appear on the right hand side of (10.44)). These particulars of the situation call for only small and obvious modifications to the numerical scheme described above.

D. An example: Disability pension with qualifying period. At time 0 an active life aged x purchases a disability insurance with term n years and conditions as follows: Premium is payable at constant rate c in active state. Pension is payable at constant rate b in disabled state, but only after an (uninterrupted) qualifying period of q years in that state. Thus the payment stream is given by

$$dB(t) = b\,1[S_t > q]\,I_i(t)\,dt \,-\, c\,I_a(t)\,dt\,,$$

where S_t is the state duration at time t.

Figure 10.4 shows a flowchart for a semi-Markov disability model apt to describe the policy. The intensities depend on the age at entry x, the policy duration t and, in invalid state, also on the state duration s. (The dependence on x does not create any difficulties – we are only interested in the intensities as function of s and t and could, in fact, have chosen not to visualize x in the notation.)

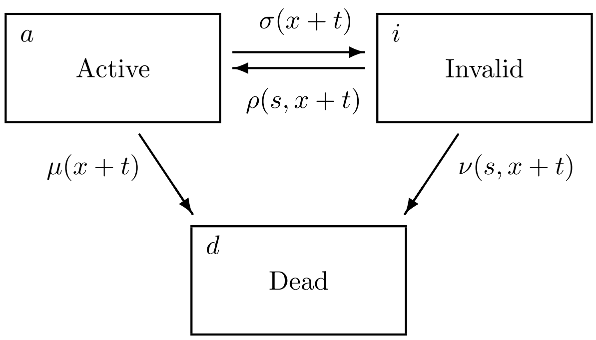


Figure 10.4: A semi-Markov chain disability model.

The force of interest at policy duration t is assumed to be deterministic and is denoted by $r(t)$.

Let $V_a(t)$ and $V_i(s,t)$ be the reserves in state a at policy duration t and the reserve in state i at state duration s and policy duration t, respectively. (Obviously, the former does not depend on state duration.) The differential equations (10.44) specialize to

$$\frac{\partial}{\partial t}V_a(t) = (r(t) + \sigma(x+t) + \mu(x+t))V_a(t) + c - \sigma(x+t)V_i(0,t)\,,$$

$$\frac{\partial}{\partial t}V_i(s,t) = (r(t)+\rho(s,x+t)+\nu(x+t))V_i(s,t) - 1[s > q]\,b - \rho(s,x+t)V_a(t) - \frac{\partial}{\partial s}V_i(s,t)\,,$$

for $0 \le s \le t \le n$, subject to the conditions

$$V_a(n-) = 0\,, \qquad V_i(s, n-) = 0\,,\ 0 \le s \le n\,.$$

The programme 'PDE-MC1.pas' is designed to compute numerical solutions to reserves in this model.

E. Maximum likelihood estimation of parameters. The likelihood function is constructed in the same way as in the Markov case. The only difference is that the state duration $S^\ell(t)$ at time t for person No. ℓ appears in the intensity functions. In the general case, with parametric intensity function $\mu_{gh}(s, t; \theta)$ of transition from state g to state h at policy duration t and state duration s, the log likelihood function based on data for m individuals is

$$\ln L = \sum_{g \ne h} \int \sum_{\ell=1}^{m} \left\{ \ln \mu_{gh}(S^\ell(t-), t; \theta)\, dN^\ell_{gh}(t) - \mu_{gh}(S^\ell(t), t, \theta) I^\ell_g(t)\, dt \right\}\,,$$

(The left-limit in the terms $\ln \mu_{gh}(S^\ell(t-), t; \theta)\, dN^\ell_{gh}(t)$ needs to be there; if there is a transition at time t, then $S^\ell(t) = 0$ by definition, whereas the jump is produced under the force of transition with state duration $S^\ell(t-)$.) The MLE equations are obtained upon differentiating w.r.t. θ and setting derivatives equal to 0:

$$\sum_{g \ne h} \int \sum_{\ell=1}^{m} \left\{ \frac{\partial}{\partial \theta_p} \ln \mu_{gh}(S^\ell(t-), t; \theta)\, dN^\ell_{gh}(t) - \frac{\partial}{\partial \theta_p} \mu_{gh}(S^\ell(t), t, \theta) I^\ell_g(t)\, dt \right\} = 0\,,$$

$p = 1, \ldots, s$. The integrals w.r.t. the counting processes are of course just sums over the times of transitions (they are empty and equal to 0 if no transitions of the sort occurred, of course):

$$\sum_{g \ne h} \sum_{\ell} \left\{ \sum_j \frac{\partial}{\partial \theta_p} \ln \mu_{gh}(S^\ell(T^{\ell j}_{gh}-), T^{\ell j}_{gh}; \theta) - \int \frac{\partial}{\partial \theta_p} \mu_{gh}(S^\ell(t), t, \theta) I^\ell_g(t)\, dt \right\} = 0\,,$$

where $T^{\ell j}_{gh}$ is the time of the $j-th$ transition from state g to state h for individual No ℓ.

Now it remains just to calculate the derivatives of the specified intensities w.r.t. all parameters, and find expressions for integrals of these derivatives, and put into the general form. Usually one will have to solve the ML equations numerically.

As an example, suppose that the insurance company keeps complete records of disabilities, recoveries and deaths for m lives that are assumed to be independent and governed by the model in Figure 10.4. Life No. ℓ entered at age x^ℓ and has been observed over a period ending by right-censoring at age y^ℓ. Assume that the intensities in the disability model in Paragraph D are parametric functions of the form

$$\begin{aligned}
\sigma(x+t) &= \sigma_1 + \sigma_2 e^{\sigma_3 (x+t)} \,, \\
\mu(x+t) &= \alpha + \beta e^{\gamma(x+t)} \,, \\
\rho(s, x+t) &= \rho_1 + \rho_2 e^{\rho_3 (x+t)} + \rho_4 e^{\rho_5 s} \,, \\
\nu(s, x+t) &= \alpha + \beta e^{\gamma(x+t)} \,,
\end{aligned}$$

We have to differentiate the log likelihood w.r.t. each of the parameters and set the derivative equal to 0, which gives 11 equations for the 11 parameters. For instance, the equation obtained by differentiating w.r.t. α is

$$\sum_{\ell; T^\ell < y^\ell - x^\ell} \frac{1}{\alpha + \beta e^{\gamma(x^\ell + T^\ell)}} - \sum_\ell \int_{x^\ell}^{y^\ell} (I_a^\ell(t) + I_i^\ell(t))\, dt = 0 \,,$$

where T^ℓ is the time of death (as active or invalid). The equation obtained by differentiating w.r.t. ρ_5 is

$$\sum_\ell \sum_j \frac{\rho_4 e^{\rho_5 S^\ell(T_{ia}^{\ell j}-)} S^\ell(T_{ia}^{\ell j}-)}{\rho_1 + \rho_2 e^{\rho_3 (x + T_{ia}^{\ell j})} + \rho_4 e^{\rho_5 S^\ell(T_{ia}^{\ell j}-)}} - \sum_\ell \int_{x^\ell}^{y^\ell} \rho_4 e^{\rho_5 S^\ell(t)} S^\ell(t) I_i^\ell(t)\, dt = 0 \,,$$

Write down the remaining equations.

Bibliography

[1] Aase, K.K. and Persson, S.-A. (1994). Pricing of unit-linked life insurance policies. *Scand. Actuarial J.*, **1994**, 26-52.

[2] Aczel J. *Lectures on Functional Equations and their Applications.* Academic Press, 1966.

[3] Andersen, P.K., Borgan, Ø., Gill, R.D., Keiding, N. (1993). *Statistical Models Based on Counting Processes*. Springer-Verlag, New York, Berlin, Heidelberg.

[4] Anderson, J.L. and Dow, J.B. (1948). *Actuarial Statistics, Vol. II: Constructions of Mortality and other Tables.* Cambridge University Press.

[5] Barlow, R.E. and Proschan, F (1981): *Statistical Theory of Reliability and Life Testing*, Holt, Reinhart and Winston Inc.

[6] Berger, A. (1939): *Mathematik der Lebensversicherung*. Verlag von Julius Springer, Vienna.

[7] Bibby J.M., Mardia, K.V., and Kent J.T. *Multivariate Analysis.* Academic Press, 1979.

[8] Björk, T., Kabanov, Y., Runggaldier, W. (1997): Bond market structures in the presence of marked point processes. *Mathematical Finance*, **7**, 211-239.

[9] Björk, T. (1998): *Arbitrage Theory in Continuous Time*, Oxford University Press.

[10] Black, F., Scholes, M. (1973): The pricing of options and corporate liabilities. *J. Polit. Economy*, **81**, 637-654.

[11] Bowers, N.L. Jr., Gerber, H.U., Hickman, J.C., and Nesbitt, C.J. (1986). Actuarial Mathematics. The Society of Actuaries. Itasca, Illinois.

[12] Cox, J., Ross, S., Rubinstein, M. (1979): Option pricing: A simplified approach. *J. of Financial Economics*, **7**, 229-263.

[13] Delbaen, F., Schachermayer, W. (1994): A general version of the fundamental theorem on asset pricing. *Mathematische Annalen*, **300**, 463-520.

[14] De Pril, N. (1989). The distributions of actuarial functions. *Mitteil. Ver. Schweiz. Vers.math.*, **89**, 173-183.

[15] De Vylder, F. and Jaumain, C. (1976). *Exposé moderne de la théorie mathématique des opérations viagère.* Office des Assureurs de Belgique, Bruxelles.

[16] Dhaene, J. (1990). Distributions in life insurance. *ASTIN Bull.* **20**, 81-92.

[17] Donald D.W.A. *Compound Interest and Annuities Certain.* Heinemann, London, 1970.

[18] Eberlein, E., Raible, S. (1999): Term structure models driven by general Lévy processes. *Mathematical Finance*, **9**, 31-53.

[19] Elliott, R.J., Kopp, P.E. (1998): *Mathematics of financial markets*, Springer-Verlag.

[20] Esary, J.D. and Proschan, F (1972): Relationships among some notions of bivariate dependence, *Annals of Mathematical Statistics*, **43**, 651-655.

[21] Föllmer, H., Sondermann, D. (1986): Hedging of non-redundant claims. In *Contributions to Mathematical Economics in Honor of Gerard Debreu*, 205-223, eds. Hildebrand, W., Mas-Collel, A., North-Holland.

[22] Gantmacher, F.R. (1959): *Matrizenrechnung II*, VEB Deutscher Verlag der Wissenschaften, Berlin.

[23] Gerber, H.U. (1995). *Life Insurance Mathematics*, 2nd edn. Springer-Verlag.

[24] Harrison, J.M., Kreps, D.M. (1979): Martingales and arbitrage in multiperiod securities markets. *J. Economic Theory*, **20**, 1979, 381-408.

[25] Harrison, J.M., Pliska, S. (1981): Martingales and stochastic integrals in the theory of continuous trading. *J. Stoch. Proc. and Appl.*, **11**, 215-260.

[26] Hoem, J.M. (1969): Markov chain models in life insurance. *Blätter Deutsch. Gesellschaft Vers.math.*, **9**, 91–107.

[27] Hoem, J.M. (1969): Purged and partial Markov chains. *Skandinavisk Aktuarietidskrift*, **52**, 147–155.

[28] Hoem, J.M. and Aalen, O.O. (1978). Actuarial values of payment streams. *Scand. Actuarial J.* **1978**, 38-47.

[29] Jordan, C.W. (1967). *Life Contingencies.* The Society of Actuaries, Chicago.

[30] Karlin, S., Taylor, H. (1975): *A first Course in Stochastic Processes*, 2nd. ed., Academic Press.

[31] Kellison S. (1970). *The theory of interest.* Richard D. Irwin, Inc. Homewood, Illinois.

[32] Merton, R.C. (1973): The theory of rational option pricing. *Bell Journal of Economics and Management Science*, **4**, 141-183.

[33] Merton, R.C. (1976): Option pricing when underlying stock returns are discontinuous. *J. Financial Economics*, **3**, 125-144.

[34] Møller, T. (1998): Risk minimizing hedging strategies for unit-linked life insurance. *ASTIN Bull.*, **28**, 17-47.

[35] Norberg, R. (1985). Lidstone in the continuous case. *Scand. Actuarial J.* **1985**, 27-32.

[36] Norberg, R. (1991). Reserves in life and pension insurance. *Scand. Actuarial J.* **1991**, 1-22.

[37] Norberg R. (1993). Identities for present values of life insurance benefits. *Scand. Actuarial J.*, 1993, p. 100–106.

[38] Norberg, R. (1995): Differential equations for moments of present values in life insurance. *Insurance: Math. & Econ.*, **17**, 171-180.

[39] Norberg, R. (1995): A time-continuous Markov chain interest model with applications to insurance. *J. Appl. Stoch. Models and Data Anal.*, 245-256.

[40] Norberg, R. (1998): Vasiček beyond the normal. *Working paper No. 152*, Laboratory of Actuarial Math., Univ. Copenhagen.

[41] Norberg, R. (1999): A theory of bonus in life insurance. *Finance and Stochastics.*, **3**, 373-390

[42] Norberg, R. (2001): On bonus and bonus prognoses in life insurance. *Scand. Actuarial J.*

[43] Pliska, S.R. (1997): *Introduction to Mathematical Finance*, Blackwell Publishers.

[44] Ramlau-Hansen, H. (1991): Distribution of surplus in life insurance. *ASTIN Bull.*, **21**, 57-71.

[45] Serfling R. (1980): *Approximation Theorems of Mathematical Statistics*, Wiley. (Paperback edition 2001.)

[46] Sverdrup, E. (1969): *Noen forsikringsmatematiske emner.* Stat. Memo. No. 1, Inst. of Math., Univ. of Oslo. (In Norwegian.)

[47] Thomas, J.W. (1995): *Numerical Partial Differential Equations: Finite Difference Methods*, Springer-Verlag.

[48] Width, E. (1986): A note on bonus theory. *Scand. Actuarial J.* **1986**, 121-126.

Appendix A

Calculus

A. Piecewise differentiable functions. Being concerned with operations in time, commencing at some initial date, we will consider functions defined on the positive real line $[0, \infty)$. Thus, let us consider a generic function $X = \{X_t\}_{t \geq 0}$ and think of X_t as the state or value of some process at time t. For the time being we take X to be real-valued.

In the present text we will work exclusively in the space of so-called *piecewise differentiable functions*. From a mathematical point of view this space is tiny since only elementary calculus is needed to move about in it. From a practical point of view it is huge since it comfortably accommodates any idea, however sophisticated, that an actuary may wish to express and analyse. It is convenient to enter this space from the outside, starting from a wider class of functions.

We first take X to be of *finite variation* (FV), which means that it is the difference between two non-decreasing, finite-valued functions. Then the left-limit $X_{t-} = \lim_{s \uparrow t} X_s$ and the right-limit $X_{t+} = \lim_{s \downarrow t} X_s$ exist for all t, and they differ on at most a countable set $\mathcal{D}(X)$ of discontinuity points of X.

We are particularly interested in FV functions X that are right-continuous (RC), that is, $X_t = \lim_{s \downarrow t} X_s$ for all t. Any probability distribution function is of this type, and any stream of payments accounted as incomes or outgoes, can reasonably be taken to be FV and, as a convention, RC. If X is RC, then $\Delta X_t = X_t - X_{t-}$, when different from 0, is the jump made by X at time t.

For our purposes it suffices to let X be of the form

$$X_t = X_0 + \int_0^t x_\tau \, d\tau + \sum_{0 < \tau \leq t} (X_\tau - X_{\tau-}) \,. \tag{A.1}$$

The integral, which may be taken to be of Riemann type, adds up the continuous increments/decrements, and the sum, which is understood to range over discontinuity times, adds up increments/decrements by jumps.

We assume, furthermore, that X is *piecewise differentiable* (PD); A property holds *piecewise* if it takes place everywhere except, possibly, at a finite number of points in every finite interval. In other words, the set of exceptional points, if not empty, must be of the form $\{t_0, t_1, \ldots\}$, with $t_0 < t_1 < \cdots$, and, in case it is infinite, $\lim_{j\to\infty} t_j = \infty$. Obviously, X is PD if both X and x are piecewise continuous. At any point $t \notin \mathcal{D} = \mathcal{D}(X) \cup \mathcal{D}(x)$ we have $\frac{d}{dt}X_t = x_t$, that is, the function X grows (or decays) continuously at rate x_t.

As a convenient notational device we shall frequently write (A.1) in differential form as

$$dX_t = x_t\,dt \,+\, X_t - X_{t-}\,. \tag{A.2}$$

A left-continuous PD function may be defined by letting the sum in (A.1) range only over the half-open interval $[0, t)$. Of course, a PD function may be neither right-continuous nor left-continuous, but such cases are of no interest to us.

B. The integral with respect to a function. Let X and Y both be PD and, moreover, let X be RC and given by (A.2). The integral over $(s, t]$ of Y with respect to X is defined as

$$\int_s^t Y_\tau\,dX_\tau = \int_s^t Y_\tau x_\tau\,d\tau + \sum_{s<\tau\le t} Y_\tau(X_\tau - X_{\tau-})\,, \tag{A.3}$$

provided that the individual terms on the right and also their sum are well defined. Considered as a function of t the integral is itself PD and RC with continuous increments $Y_t x_t\,dt$ and jumps $Y_t(X_t - X_{t-})$. One may think of the integral as the weighted sum of the Y-values, with the increments of X as weights, or vice versa. In particular, (A.1) can be written simply as

$$X_t = X_s + \int_s^t dX_\tau\,, \tag{A.4}$$

saying that the value of X at time t is its value at time s plus all its increments in $(s, t]$.

By definition,

$$\int_s^{t-} Y_\tau\,dX_\tau = \lim_{r\nearrow t}\int_s^r Y_\tau\,dX_\tau = \int_s^t Y_\tau\,dX_\tau - Y_t(X_t - X_{t-}) = \int_{(s,t)} Y_\tau\,dX_\tau\,,$$

a left-continuous function of t. Likewise,

$$\int_{s-}^t Y_\tau\,dX_\tau = \lim_{r\nearrow s}\int_r^t Y_\tau\,dX_\tau = \int_s^t Y_\tau\,dX_\tau + Y_s(X_s - X_{s-}) = \int_{[s,t]} Y_\tau\,dX_\tau\,,$$

a left-continuous function of s.

C. The chain rule (Itô's formula). Let $X_t = (X_t^1, \ldots, X_t^m)$ be an m-variate function with PD and RC components given by $dX_t^i = x_t^i\,dt + (X_t^i - X_{t-}^i)$. Let $f : \mathcal{R}^m \mapsto \mathcal{R}$ have continuous partial derivatives, and form the composed function $f(X_t)$. On the open intervals where there are neither discontinuities in the x^i nor jumps of the X^i, the function $f(X_t)$ develops in accordance with the well-known chain rule for scalar fields along rectifiable curves. At the exceptional points $f(X_t)$ may change (only) due to jumps of the X^i, and at any such point t it jumps by $f(X_t) - f(X_{t-})$. Thus, we gather the so-called *change of variable rule* or *Itô's formula*, which in our simple function space reads

$$df(X_t) = \sum_{i=1}^{m} \frac{\partial f}{\partial x^i}(X_t)\; x_t^i\, dt + f(X_t) - f(X_{t-})\,, \tag{A.5}$$

or, in integral form,

$$f(X_t) = f(X_s) + \int_s^t \sum_{i=1}^{m} \frac{\partial}{\partial x^i} f(X_\tau)\, x_\tau^i\, d\tau + \sum_{s<\tau\le t} \{f(X_\tau) - f(X_{\tau-})\}\,. \tag{A.6}$$

Obviously, $f(X_t)$ is PD and RC.

A frequently used special case is (check the formulas!)

$$\begin{aligned} d(X_tY_t) &= X_t\, y_t\, dt + Y_t\, x_t\, dt + X_tY_t - X_{t-}Y_{t-} \\ &= X_{t-}\, dY_t + Y_{t-}\, dX_t + (X_t - X_{t-})(Y_t - Y_{t-}) \\ &= X_{t-}\, dY_t + Y_t\, dX_t\,. \end{aligned} \tag{A.7}$$

If X and Y have no common jumps, as is certainly the case if one of them is continuous, then (A.7) reduces to the familiar

$$d(X_tY_t) = X_t\, dY_t + Y_t\, dX_t\,. \tag{A.8}$$

The integral form of (A.7) is the so-called rule of *integration by parts*:

$$\int_s^t Y_\tau dX_\tau = Y_tX_t - Y_sX_s - \int_s^t X_{\tau-}\, dY_\tau\,. \tag{A.9}$$

Let us consider three special cases for which (A.9) can be obtained by direct calculation and specialises to well-known formulas. Setting $s = 0$ (just a matter of notation), (A.9) can be cast as

$$Y_tX_t = Y_0X_0 + \int_0^t Y_\tau\, dX_\tau + \int_0^t X_{\tau-}\, dY_\tau\,, \tag{A.10}$$

which shows how the product of X and Y at time t emerges from its initial value at time 0 plus all its increments in the interval $(0, t]$.

Assume first that X and Y are both discrete. To keep notation simple assume $X_t = \sum_{j=0}^{[t]} x_j$ and $Y_t = \sum_{j=0}^{[t]} y_j$. Then

$$X_tY_t = \sum_{i=0}^{[t]} x_i \sum_{j=0}^{[t]} y_j$$

$$
\begin{aligned}
&= x_0y_0 + \sum_{i=1}^{[t]}\sum_{j=0}^{i} y_j\, x_i + \sum_{j=1}^{[t]}\sum_{i=0}^{j-1} x_i\, y_j \\
&= X_0Y_0 + \sum_{i=1}^{[t]} Y_i\, x_i + \sum_{j=1}^{[t]} X_{j-1}\, y_j \\
&= X_0\, Y_0 + \int_0^t Y_\tau\, dX_\tau + \int_0^t X_{\tau-}\, dY_\tau\,,
\end{aligned}
$$

which is (A.10). We see here that the left limit on the right of (A.10) is essential. This case is basically nothing but the rule of changing the order of summation in a double sum, the only new thing being that we formally consider the sums X and Y as functions of a continuous time index; only the values at integer times matter, however.

Assume next that X and Y are both continuous, that is,

$$X_t = X_0 + \int_0^t x_\tau\, d\tau\,, Y_t = Y_0 + \int_0^t y_\tau\, d\tau\,.$$

Take $X_0 = Y_0 = 0$ for the time being. Then

$$
\begin{aligned}
X_tY_t &= \int_0^t x_\sigma\, d\sigma \int_0^t y_\tau\, d\tau \\
&= \int\int_{0<\tau\le\sigma\le t} y_\tau\, d\tau\, x_\sigma\, d\sigma + \int\int_{0<\sigma<\tau\le t} x_\sigma\, d\sigma\, y_\tau\, d\tau \\
&= \int_0^t \int_0^\sigma y_\tau\, d\tau\, x_\sigma\, d\sigma + \int_0^t \int_0^{\tau-} x_\sigma\, d\sigma\, y_\tau\, d\tau \\
&= \int_0^t Y_\sigma x_\sigma\, d\sigma + \int_0^t X_{\tau-}\, y_\tau\, d\tau \\
&= \int_0^t Y_\tau\, dX_\tau + \int_0^t X_\sigma\, dY_\sigma\,,
\end{aligned}
$$

which also conforms with (A.10): the left limit in the next to last line disappeared since an integral with respect to $d\tau$ remains unchanged if we change the integrand at a countable set of points. The result for general X_0 and Y_0 is obtained by applying the formula above to $X_t - X_0$ and $Y_t - Y_0$.

Finally, let one function be discrete and the other continuous, e.g. $X_t = \sum_{j=0}^{[t]} x_j$ and $Y_t = Y_0 + \int_0^t y_\tau\, d\tau$. Introduce

$$\hat{y}_0 = Y_0\,, \qquad \hat{y}_j = \int_{j-1}^{j} y_\tau\, d\tau\,,\ j = 1, \ldots, [t].$$

We have $X_t = X_{[t]}$ and

$$X_tY_t \;=\; X_{[t]}Y_{[t]} + X_{[t]}(Y_t - Y_{[t]})$$

$$= \sum_{i=0}^{[t]} x_i \sum_{j=0}^{[t]} \hat{y}_j + X_{[t]} \int_{[t]}^{t} y_\tau \, d\tau \,. \tag{A.11}$$

Upon applying our first result for two discrete functions, the first term in (A.11) becomes

$$\begin{aligned} & X_0 Y_0 + \sum_{i=1}^{[t]} \sum_{j=0}^{i} \hat{y}_j \, x_i + \sum_{j=1}^{[t]} \sum_{i=0}^{j-1} x_i \, \hat{y}_j \\ &= X_0 Y_0 + \sum_{i=1}^{[t]} Y_j \, x_i + \sum_{j=1}^{[t]} X_{j-1} \int_{j-1}^{j} y_\tau \, d\tau \\ &= X_0 Y_0 + \int_0^t Y_\tau \, dX_\tau + \int_0^{[t]} X_{\tau-} dY_\tau \end{aligned}$$

The second term in (A.11) is $\int_{[t]}^t X_{\tau-} dY_\tau$. Thus, also in this case we arrive at (A.10). Again the left-limit is irrelevant since $dY_\tau = y_\tau \, d\tau$.

The general formula now follows from these three special cases by the fact that the the integral is a linear operator with respect to the integrand and the integrator.

D. Counting processes. Let $t_1 < t_2 < \cdots$ be a sequence in $(0, \infty)$, either finite or, if infinite, such that $\lim_{j\to\infty} t_j = \infty$. Think of t_j as the j-th time of occurrence of a certain event. The number of events occurring within a given time t is $N_t = \sharp\{j\,;\, t_j \le t\} = \sum_j 1_{[t_j,\infty)}(t)$ or, putting $t_0 = 0$, $N_t = j$ for $t_j \le t < t_{j+1}$. The function $N = \{N_t\}_{t\ge 0}$ thus defined is called a *counting function* since it currently counts the number of occurred events. It is a particularly simple PD and RC function commencing from $N_0 = 0$ and thereafter increasing only by jumps of size 1 at the epochs t_j, $j = 1, 2, \ldots$

The change of variable rule (A.6) becomes particularly simple when X is a counting function. In fact, for $f : \mathcal{R} \mapsto \mathcal{R}$ and for N defined above,

$$\begin{aligned} f(N_t) &= f(N_s) + \sum_{s<\tau\le t} \{f(N_\tau) - f(N_{\tau-})\} && \text{(A.12)} \\ &= f(N_s) + \sum_{s<\tau\le t} \{f(N_{\tau-}+1) - f(N_{\tau-})\}(N_\tau - N_{\tau-}) && \text{(A.13)} \\ &= f(N_s) + \int_s^t \{f(N_{\tau-}+1) - f(N_{\tau-})\} \, dN_\tau \,. && \text{(A.14)} \end{aligned}$$

Basically, what these expressions state, is just

$$f(j) = f(0) + \sum_{i=1}^{j} \{f(i) - f(i-1)\} \,.$$

Still they will prove to be useful representations when we come to stochastic counting processes.

Going back to the general PD and RC function X in (A.1), we can associate with it a counting function N defined by $N_t = \sharp\{\tau \in (0,t];\, X_\tau \neq X_{\tau-}\}$, the number of discontinuities of X within time t. Equipped with our notion of integral, we can now express X as

$$dX_t = x_t^c\, dt + x_t^d\, dN_t\,, \tag{A.15}$$

where $x_t^c = x_t$ is the instantaneous rate of continuous change and $x_t^d = X_t - X_{t-}$ is the size of the jump, if any, at t. Generalizing (A.14), we have

$$f(X_t) = f(X_s) + \int_s^t \frac{d}{dx} f(X_\tau)\, x_\tau^c\, d\tau + \int_s^t \{f(X_{\tau-} + x_\tau^d) - f(X_{\tau-})\}\, dN_\tau\,. \tag{A.16}$$

Appendix B

Indicator functions

A. Indicator functions in general spaces. Let Ω be some space with generic point ω, and let A be some subset of Ω. The function $I_A : \Omega \to \{0, 1\}$ defined by

$$I_A(\omega) = \begin{cases} 1 & \text{if} \quad \omega \in A\,, \\ 0 & \text{if} \quad \omega \in A^c\,, \end{cases}$$

is called the *indicator function* or just the *indicator* of A since it indicates by the value 1 precisely those points ω that belong to A.

Since I_A assumes only the values 0 and 1, $(I_A)^p = I_A$ for any $p > 0$. Clearly, $I_\emptyset = 0$, $I_\Omega = 1$, and

$$I_{A^c} = 1 - I_A\,, \tag{B.1}$$

where $A^c = \Omega \backslash A$ is the complement of A.

For any two sets A and B (subsets of Ω),

$$I_{A \cap B} = I_A I_B \tag{B.2}$$

and

$$I_{A \cup B} = I_A + I_B - I_A I_B\,. \tag{B.3}$$

The last two statements are displayed here only for ease of reference. They are special cases of the following results, valid for any finite collection of sets $\{A_1, \ldots, A_r\}$:

$$I_{\cap_{j=1}^r A_j} = \prod_{j=1}^{r} I_{A_j}\,, \tag{B.4}$$

$$I_{\cup_{j=1}^r A_j} = \sum_j I_{A_j} - \sum_{j_1 < j_2} I_{A_{j_1}} I_{A_{j_2}} + \ldots + (-1)^{r-1} I_{A_1} \cdots I_{A_r}\,. \tag{B.5}$$

The relation (B.4) is obvious. To demonstrate (B.5), we need the identity

$$\prod_{j=1}^{r} (a_j + b_j) \quad = \quad \sum_{p=0}^{r} \sum_{r \backslash p} a_{j_1} \cdots a_{j_p} b_{j_{p+1}} \cdots b_{j_r}\,, \tag{B.6}$$

where $r\backslash p$ signifies that the sum ranges over all $\binom{r}{p}$ different ways of dividing $\{1,\ldots,r\}$ into two disjoint subsets $\{j_1,\ldots,j_p\}$ ($\emptyset$ when $p=0$) and $\{j_{p+1},\ldots,j_r\}$ ($\emptyset$ when $p=r$). Combining the general relation

$$\{\cup_\alpha A_\alpha\}^c = \cap_\alpha A_\alpha^c\,, \tag{B.7}$$

with (B.1) and (B.4), we find

$$I_{\cup_{j=1}^r A_j} = 1 - I_{\cap_{j=1}^r A_j^c} = 1 - \prod_{j=1}^r (1 - I_{A_j})\,,$$

and arrive at (B.5) by use of (B.6).

B. Further aspects of indicators. The algebraic expressions in (B.4) and (B.5) apply only to the finite case. For any collection $\{A_\alpha\}$ of sets indexed by α ranging in an arbitrary space, possibly uncountable,

$$I_{\cap_\alpha A_\alpha} = \inf_\alpha I_{A_\alpha} \tag{B.8}$$

and

$$I_{\cup_\alpha A_\alpha} = \sup_\alpha I_{A_\alpha}\,. \tag{B.9}$$

In fact, inf and sup are attained here, so we can write min and max. In accordance with the latter two results one may define $\sup_\alpha A_\alpha = \cup_\alpha A_\alpha$ and $\inf_\alpha A_\alpha = \cap_\alpha A_\alpha$.

The relation (B.7) rests on elementary logical operations, but also follows from $1 - \sup_\alpha I_{A_\alpha} = \inf_\alpha (1 - I_{A_\alpha})$.

The representation of sets by indicators supports the understanding of some conventions and definitions in set theory. For instance, if $\{A_j\}_{j=1,2,\ldots}$ is a disjoint sequence of sets, some authors write $\sum_j A_j$ instead of $\cup_j A_j$. This is motivated by

$$I_{\cup_j A_j} = \sum_j I_{A_j}\,,$$

valid for disjoint sets.

For any sequence $\{A_j\}$ of sets one writes $\limsup A_j$ for the set of points ω that belong to infinitely many of the A_j, that is,

$$\limsup A_j = \cap_j \cup_{k\geq j} A_k$$

(for all j there exists some $k \geq j$ such that ω belongs to A_k). By $\liminf A_j$ is meant the set of points ω that belong to all but possibly finitely many of the A_j, that is,

$$\liminf A_j = \cup_j \cap_{k\geq j} A_k$$

(there exists a j such that for all $k \geq j$ the point ω belongs to A_k). This usage is in accordance with

$$I_{\cap_j \cup_{k\geq j} A_k} = \inf_j \sup_{k\geq j} I_{A_k}$$

and

$$I_{\cup_j \cap_{k \geq j} A_k} = \sup_j \inf_{k \geq j} I_{A_k} \,,$$

obtained upon combining (B.8) and (B.9).

C. Indicators of events. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be some probability space. The indicator I_A of an event $A \in \mathcal{F}$ is a simple binomial random variable;

$$I_A \sim \mathrm{Bin}(1, \mathbb{P}[A]) \,.$$

It follows that

$$\mathbb{E}\,[I_A] = \mathbb{P}[A] \,, \;\; \mathbb{V}\,[I_A] = \mathbb{P}[A](1 - \mathbb{P}[A]) \,. \tag{B.10}$$

Since we often will need to equip indicator functions with subscripts, we will use the notation $1[A]$ and I_A interchangeably.

Appendix C

Distribution of the number of occurring events

A. The main result. Let $\{A_1, \ldots, A_r\}$ be a finite assembly of events, not necessarily disjoint. Introduce the short-hand $I_j = I_{A_j}$. We seek the probability distribution of the number of events that occur out of the total of r events,

$$Q = \sum_{j=1}^{r} I_j \, .$$

It turns out that this distribution can be expressed in terms of the probabilities of intersections of selections from the assembly of sets. Introduce

$$Z_p = \sum_{j_1 < \ldots < j_p} \mathbb{P}\left[A_{j_1} \cap \cdots \cap A_{j_p}\right] \, , \;\; p = 1, \ldots, r, \tag{C.1}$$

and define in particular $Z_0 = 1$.

Theorem
The probability distribution of Q can be expressed by the Z_p in (C.1) as

$$\mathbb{P}[Q = q] \;\; = \;\; \sum_{p=q}^{r} (-1)^{p-q} \binom{p}{p-q} Z_p \, , \quad q = 0, \ldots, r \, , \tag{C.2}$$

$$\mathbb{P}[Q \geq q] \;\; = \;\; \sum_{p=q}^{r} (-1)^{p-q} \binom{p-1}{p-q} Z_p \, , \quad q = 1, \ldots, r \, . \tag{C.3}$$

Proof: Obviously,

$$\{Q = q\} = \bigcup_{r \setminus q} A_{j_1} \cap \cdots \cap A_{j_q} \cap A^c_{j_{q+1}} \cap \cdots \cap A^c_{j_r} \, .$$

The elements in the union are mutually disjoint, and so

$$I_{\{Q=q\}} = \sum_{r\backslash q} I_{j_1} \cdots I_{j_q}(1 - I_{j_{q+1}}) \cdots (1 - I_{j_r})\,.$$

Starting from this expression, the generating function of the sequence $\{I_{\{Q=p\}}\}_{p=0,\ldots,r}$ can be shaped as follows by repeated use (B.6):

$$\begin{aligned}
\sum_{p=0}^{r} s^p I_{\{Q=p\}} &= \sum_{p=0}^{r} s^p \sum_{r\backslash p} I_{j_1} \cdots I_{j_p}(1 - I_{j_{p+1}}) \cdots (1 - I_{j_r}) \\
&= \sum_{p=0}^{r} \sum_{r\backslash p} sI_{j_1} \cdots sI_{j_p}(1 - I_{j_{p+1}}) \cdots (1 - I_{j_r})\,. \\
&= \prod_{j=1}^{r} (sI_j + 1 - I_j) \\
&= \prod_{j=1}^{r} ((s-1)I_j + 1) \\
&= \sum_{p=0}^{r} \sum_{r\backslash p} (s-1)I_{j_1} \cdots (s-1)I_{j_p}\, 1^{r-p}\,,
\end{aligned}$$

where the first term corresponding to $p = 0$ is to be interpreted as 1. Thus,

$$\sum_{p=0}^{r} s^p I_{\{Q=p\}} = \sum_{p=0}^{r} (s-1)^p Y_p\,, \tag{C.4}$$

where

$$Y_p = \sum_{j_1 < \ldots < j_p} I_{j_1} \cdots I_{j_p}\,, \quad p = 1, \ldots, r, \tag{C.5}$$

and $Y_0 = 1$. Upon differentiating (C.4) q times with respect to s and putting $s = 0$, we get

$$q!I_{\{Q=q\}} = \sum_{p=q}^{r} p^{(q)}(-1)^{p-q} Y_p\,,$$

hence, noting that $p^{(q)}/q! = \binom{p}{q} = \binom{p}{p-q}$,

$$I_{\{Q=q\}} = \sum_{p=q}^{r} (-1)^{p-q} \binom{p}{p-q} Y_p\,. \tag{C.6}$$

Taking expectation, we arrive at (C.2).

To prove (C.3), insert $I_{\{Q=p\}} = I_{\{Q\geq p\}} - I_{\{Q\geq p+1\}}$ on the left of (C.4) and rearrange as follows:

$$\begin{aligned}\sum_{p=0}^{r} s^p \left(I_{\{Q\geq p\}} - I_{\{Q\geq p+1\}}\right) &= \sum_{p=0}^{r} s^p I_{\{Q\geq p\}} - \sum_{p=1}^{r} s^{p-1} I_{\{Q\geq p\}} \\ &= 1 + \sum_{p=1}^{r}(s-1)s^{p-1} I_{\{Q\geq p\}}\,.\end{aligned}$$

Thus, recalling that $Y_0 = 1$, (C.4) is equivalent to

$$\sum_{p=1}^{r} s^{p-1} I_{\{Q\geq p\}} = \sum_{p=1}^{r}(s-1)^{p-1} Y_p\,.$$

Differentiating here $q-1$ times with respect to s and putting $s=0$, gives

$$(q-1)!I_{\{Q\geq q\}} = \sum_{p=q}^{r}(p-1)^{(q-1)}(-1)^{p-q} Y_p\,,$$

hence

$$I_{\{Q\geq q\}} = \sum_{p=q}^{r}(-1)^{p-q}\binom{p-1}{p-q} Y_p\,, \tag{C.7}$$

which implies (C.3). $\square$

B. Comments and examples. Setting all A_j equal to the sure event Ω, all the indicators I_j become identically 1. Thus Y_p defined by (C.5) becomes $\binom{r}{p}$, and (C.7) specializes to

$$1 = \sum_{p=q}^{r}(-1)^{p-q}\binom{p-1}{p-q}\binom{r}{p}\,. \tag{C.8}$$

For $q = r$, the theorem reduces to the trivial result

$$P[Q=r] = \mathbb{P}[Q\geq r] = \mathbb{P}\left[A_1\cap\cdots\cap A_r\right]\,.$$

Putting $q=1$ in (C.3) and noting that $[Q\geq 1] = \cup_j A_j$, yields

$$\begin{aligned}\mathbb{P}\left[\cup_j A_j\right] &= \sum_j \mathbb{P}[A_j] - \sum_{j_1<j_2}\mathbb{P}[A_{j_1}\cap A_{j_2}] \\ &\quad + \ldots + (-1)^{r-1}\mathbb{P}[A_1\cap\cdots\cap A_r]\,,\end{aligned}$$

which also results upon taking expectation in (B.5). This is the well-known general addition rule for probabilities, called so because it generalizes the elementary rule $\mathbb{P}[A\cup B] = \mathbb{P}[A] + \mathbb{P}[B] - \mathbb{P}[A\cap B]$ (see (B.3)). The theorem states the most general results of this type.

As a non-standard example, let us find the probability of exactly two occurrences among three events, A_1, A_2, A_3. Putting $r = 3$ and $q = 2$ in (C.2), gives

$$\mathbb{P}[Q = 2] = \mathbb{P}[A_1 \cap A_2] + \mathbb{P}[A_1 \cap A_3] + \mathbb{P}[A_2 \cap A_3] - 3\mathbb{P}[A_1 \cap A_2 \cap A_3]\,.$$

From (C.3) we obtain the probability of at least two occurrences,

$$\mathbb{P}[Q \geq 2] = \mathbb{P}[A_1 \cap A_2] + \mathbb{P}[A_1 \cap A_3] + \mathbb{P}[A_1 \cap A_3] - 2\mathbb{P}[A_1 \cap A_2 \cap A_3]\,.$$

The usefulness of the theorem is due to the decomposition of complex events into more elementary ones. The observant reader may have asked why intersections rather than unions are taken as the elementary events. The reason for doing so is apparent in the case of independent events, since then $\mathbb{P}\left[\cap_{i=1}^{p} A_{j_i}\right] = \prod_{i=1}^{p} \mathbb{P}[A_{j_i}]$, and the expressions in (C.1) – (C.3) can be computed from the $\mathbb{P}[A_j]$ by elementary algebraic operations.

Note that the results in the theorem are independent of the probability measure involved; they rest entirely on the set-relations (C.6) and (C.7).

Appendix D

Asymptotic results from statistics

A. The central limit theorem Let $\mathbf{X}_1, \mathbf{X}_2, \ldots$ be a sequence of square integrable random $q \times 1$ vectors (i.e. their variances are finite). Denote

$$\boldsymbol{\xi}_i = \mathbb{E}[\mathbf{X}_i]\,, \quad \boldsymbol{\Sigma}_i = \mathbb{V}[\mathbf{X}_i]\,,\ i = 1, 2, \ldots, \tag{D.1}$$

and introduce

$$\mathbf{B}_n = \sum_{i=1}^{n} \boldsymbol{\Sigma}_i\ ,\ n = 1, 2, \ldots \tag{D.2}$$

The celebrated Lindeberg/Feller central limit theorem states that if

$$\mathbf{a}'\mathbf{B}_n\mathbf{a} \to \infty \quad \text{and} \quad \frac{\sum_{i=1}^{n} \mathbb{E}\left[(\mathbf{a}'\mathbf{X}_i)^2 1[(\mathbf{a}'\mathbf{X}_i)^2 > \varepsilon \mathbf{a}'\mathbf{B}_n\mathbf{a}]\right]}{\mathbf{a}'\mathbf{B}_n\mathbf{a}} \to 0\,,\ \forall \mathbf{a} \neq \mathbf{0}\,,\ \forall \varepsilon > 0, \tag{D.3}$$

as $n \to \infty$, then

$$\mathbf{B}_n^{-\frac{1}{2}} \sum_{i=1}^{n} (\mathbf{X}_i - \boldsymbol{\xi}_i) \xrightarrow{L} \mathrm{N}(\mathbf{0}, \mathbf{I})\,. \tag{D.4}$$

B. Asymptotic properties of MLE estimators The asymptotic distributions derived in Chapter 8 could be obtained directly from the following standard result, which is cited here without proof.

Let $X_1, X_2, \ldots$ be a sequence of random elements with joint distribution depending on a parameter θ that varies in an open set in the s-dimensional euclidean space. Assume that the likelihood function of $X_1, X_2, \ldots, X_n$, denoted by $\Lambda_n(X_1, X_2, \ldots, X_n, \theta)$, is twice continuously differentiable with respect to θ and that the equation

$$\frac{\partial}{\partial\theta} \ln \Lambda_n(X_1, X_2, \ldots, X_n, \theta) = 0^{s\times 1}$$

has a unique solution $\hat{\theta}_n(X_1, X_2, \dots, X_n)$, called the MLE (maximum likelihood estimator). Then, if the matrix

$$\Sigma(\theta) = \left(-\mathbb{E} \left[\frac{\partial^2}{\partial\theta\partial\theta'} \ln \Lambda_n \right] \right)^{-1} \tag{D.5}$$

tends to $0^{s \times s}$ as $n \to \infty$, the MLE is asymptotically normally distributed,

$$\hat{\theta} \sim_{\mathrm{as}} \mathrm{N}(\theta, \Sigma(\theta)) \, .$$

C. The delta method Assume that $\hat{\theta}$ is a consistent estimator of $\theta \in \Theta$, an open set in $\mathcal{R}^s$, and that $\hat{\theta} \sim_{\mathrm{as}} \mathrm{N}(\theta, \Sigma)$. If $g : \mathcal{R}^s \to \mathcal{R}^r$ is a twice continuously differentiable function of θ, then

$$g(\hat{\theta}) \sim_{\mathrm{as}} \mathrm{N} \left(g(\theta), \, \frac{\partial}{\partial\theta'} g(\theta) \Sigma \frac{\partial}{\partial\theta} g(\theta) \right) \, . \tag{D.6}$$

The result follows easily by inspection of the first order Taylor expansion of $g(\hat{\theta})$ around θ.

Appendix E

The G82M mortality table

Table E.1: The mortality table G82M

x	μ_x	$f(x)$	$\ell_x = 10^5 \bar{F}(x)$	d_x	q_x
0	0.00057586	0.00057586	100 000	58	0.00057911
1	0.00058279	0.00058246	99 942	59	0.00058635
2	0.00059036	0.00058968	99 883	59	0.00059426
3	0.00059863	0.00059758	99 824	60	0.00060289
4	0.00060765	0.00060621	99 764	61	0.00061231
5	0.00061749	0.00061565	99 703	62	0.00062259
6	0.00062823	0.00062598	99 641	63	0.00063381
7	0.00063996	0.00063726	99 578	65	0.00064606
8	0.00065276	0.00064958	99 513	65	0.00065942
9	0.00066672	0.00066304	99 448	67	0.00067401
10	0.00068197	0.00067775	99 381	69	0.00068993
11	0.00069861	0.00069380	99 312	70	0.00070731
12	0.00071677	0.00071134	99 242	72	0.00072627
13	0.00073659	0.00073048	99 170	74	0.00074697
14	0.00075823	0.00075137	99 096	77	0.00076956
15	0.00078184	0.00077417	99 019	78	0.00079422
16	0.00080761	0.00079906	98 941	81	0.00082113
17	0.00083574	0.00082621	98 860	85	0.00085050
18	0.00086644	0.00085583	98 775	87	0.00088256
19	0.00089994	0.00088814	98 688	90	0.00091754

x	μ_x	$f(x)$	$\ell_x = 10^8 \bar{F}(x)$	d_x	q_x
20	0.00093652	0.00092338	98 598	94	0.00095573
21	0.00097643	0.00096182	98 504	99	0.00099740
22	0.00102000	0.00100373	98 405	102	0.00104288
23	0.00106754	0.00104942	98 303	108	0.00109252
24	0.00111944	0.00109924	98 195	112	0.00114669
25	0.00117608	0.00115353	98 083	119	0.00120582
26	0.00123790	0.00121270	97 964	124	0.00127034
27	0.00130538	0.00127718	97 840	131	0.00134076
28	0.00137902	0.00134743	97 709	139	0.00141762
29	0.00145940	0.00142394	97 570	146	0.00150150
30	0.00154713	0.00150727	97 424	155	0.00159304
31	0.00164288	0.00159800	97 269	165	0.00169293
32	0.00174738	0.00169678	97 104	175	0.00180196
33	0.00186144	0.00180428	96 929	186	0.00192094
34	0.00198594	0.00192125	96 743	199	0.00205078
35	0.00212181	0.00204849	96 544	211	0.00219247
36	0.00227011	0.00218686	96 333	226	0.00234710
37	0.00243197	0.00233728	96 107	242	0.00251584
38	0.00260863	0.00250075	95 865	259	0.00269998
39	0.00280144	0.00267834	95 606	277	0.00290091
40	0.00301189	0.00287119	95 329	298	0.00312018
41	0.00324157	0.00308050	95 031	319	0.00335944
42	0.00349226	0.00330759	94 712	343	0.00362051
43	0.00376588	0.00355382	94 369	369	0.00390537
44	0.00406451	0.00382066	94 000	396	0.00421619
45	0.00439045	0.00410964	93 604	426	0.00455532
46	0.00474620	0.00442239	93 178	459	0.00492533
47	0.00513447	0.00476062	92 719	494	0.00532902
48	0.00555825	0.00512607	92 225	532	0.00576943
49	0.00602077	0.00552060	91 693	574	0.00624989
50	0.00652560	0.00594609	91 119	617	0.00677402
51	0.00707658	0.00640446	90 502	665	0.00734576
52	0.00767794	0.00689767	89 837	716	0.00796941
53	0.00833430	0.00742765	89 121	770	0.00864963
54	0.00905067	0.00799632	88 351	830	0.00939153
55	0.00983254	0.00860553	87 521	893	0.01020062
56	0.01068591	0.00925700	86 628	960	0.01108295
57	0.01161732	0.00995232	85 668	1 032	0.01204506
58	0.01263389	0.01069283	84 636	1 108	0.01309408
59	0.01374341	0.01147959	83 528	1 189	0.01423775

x	μ_x	$f(x)$	$\ell_x = 10^8 \bar{F}(x)$	d_x	q_x
60	0.01495440	0.01231325	82 339	1 275	0.01548448
61	0.01627611	0.01319402	81 064	1 366	0.01684342
62	0.01771868	0.01412149	79 698	1 460	0.01832447
63	0.01929317	0.01509456	78 238	1 560	0.01993841
64	0.02101162	0.01611127	76 678	1 664	0.02169690
65	0.02288721	0.01716867	75 014	1 771	0.02361259
66	0.02493430	0.01826263	73 243	1 882	0.02569916
67	0.02716858	0.01938769	71 361	1 996	0.02797145
68	0.02960717	0.02053690	69 365	2 112	0.03044546
69	0.03226874	0.02170163	67 253	2 229	0.03313851
70	0.03517368	0.02287138	65 024	2 345	0.03606928
71	0.03834425	0.02403370	62 679	2 461	0.03925790
72	0.04180475	0.02517403	60 218	2 573	0.04272605
73	0.04558167	0.02627566	57 645	2 680	0.04649704
74	0.04970395	0.02731973	54 965	2 781	0.05059590
75	0.05420317	0.02828534	52 184	2 873	0.05504946
76	0.05911381	0.02914974	49 311	2 953	0.05988641
77	0.06447348	0.02988871	46 358	3 020	0.06513740
78	0.07032323	0.03047703	43 338	3 069	0.07083506
79	0.07670789	0.03088920	40 269	3 102	0.07701411
80	0.08367637	0.03110030	37 167	3 111	0.08371127
81	0.09128204	0.03108704	34 056	3 098	0.09096537
82	0.09958318	0.03082907	30 958	3 059	0.09881727
83	0.10864338	0.03031033	27 899	2 994	0.10730975
84	0.11853205	0.02952051	24 905	2 901	0.11648747
85	0.12932494	0.02845661	22 004	2 781	0.12639675
86	0.14110474	0.02712418	19 223	2 635	0.13708534
87	0.15396168	0.02553851	16 588	2 465	0.14860208
88	0.16799427	0.02372520	14 123	2 274	0.16099656
89	0.18331000	0.02172028	11 849	2 066	0.17431855
90	0.20002621	0.01956945	9 783	1 845	0.18861740
91	0.21827096	0.01732660	7 938	1 619	0.20394126
92	0.23818401	0.01505134	6 319	1 392	0.22033622
93	0.25991791	0.01280578	4 927	1 172	0.23784520
94	0.28363917	0.01065073	3 755	963	0.25650675
95	0.30952951	0.00864156	2 792	772	0.27635358
96	0.33778728	0.00682433	2 020	601	0.29741104
97	0.36862894	0.00523248	1 419	453	0.31969526
98	0.40229077	0.00388474	966	332	0.34321126
99	0.43903065	0.00278447	634	233	0.36795078

x	μ_x	$f(x)$	$\ell_x = 10^8 \bar{F}(x)$	d_x	q_x
100	0.47913004	0.00192066	401	158	0.39389013
101	0.52289614	0.00127047	243	102	0.42098791
102	0.57066422	0.00080282	141	64	0.44918271
103	0.62280022	0.00048261	77	37	0.47839101
104	0.67970357	0.00027473	40	20	0.50850528
105	0.74181017	0.00014737	20	11	0.53939240
106	0.80959582	0.00007408	9	5	0.57089277
107	0.88357981	0.00003469	4	2	0.60281998
108	0.96432893	0.00001504	2	1	0.63496163
109	1.05246177	0.00000599	1	1	0.66708109
110	1.14865351	0.00000218	0	0	0.69892078