Convex order and multistate life insurance contracts

by

J. Spreeuw

Actuarial Research Paper No. 129

December 2000

ISBN 1 901615 50 2
“Any opinions expressed in this paper are my/our own and not necessarily those of my/our employer or anyone else I/we have discussed them with. You must not copy this paper or quote it without my/our permission.”
Convex order and multistate life insurance contracts

Jaap Spreeuw

Abstract

The distribution function of the present value of a cash flow can be approximated by means of a distribution function of a random variable, which is also the present value of a sequence of payments, but has a simpler structure. The corresponding random variable has the same expectation as the random variable corresponding to the original distribution function and is a stochastic upper bound of convex order. In this paper it will be shown that such an approach can also be adopted for some multistate life insurance contracts under Markov assumptions. The quality of the approximation will be investigated by comparing the distribution obtained with the one derived from the algorithm presented in the paper by Hesselager & Norberg (1996).

Keywords: Convex order, comonotonic joint distribution, multistate life insurance contracts, present value distributions.

1 Introduction

In life contingencies under a stochastic framework, distributions of the present value of future payments are a key component in order to derive premiums satisfying a certain criterion. The most usual principle is that of actuarial equivalence, meaning that premiums are such that the expected present value of benefits less premiums is equal to zero.

The probability distribution of present values gives an indication of the riskiness of a contract, such as the variability of the actual benefits paid out or the upper tail of the present values.

De Pril (1989) and Dhaene (1990) derive distributions of such present values in a classical life insurance framework where only the two states "Alive" and "Dead" are relevant (and where only payment by single premium is considered). As the present value depends on the outcome of only one random variable, such analyses turn out to be quite straightforward.

Deriving distributions such as above in general tends to be much more complicated if instead we are dealing with life contracts involving more than two states. Hesselager & Norberg (1996) show that an approximated distribution for a multistate life insurance contract under the Markov assumption can be obtained by deriving a set of integral equations. However, the execution of the algorithm they derive may be time-consuming, especially in case there are many states and/or if the Markov chain is not hierarchical. The question is whether a method exists to derive an
approximate version of the real present value distribution, which saves computing time (though maybe at the cost of some accuracy).

A common method is to replace a random variable by a "riskier" one, i.e. a random variable being larger with respect to some ordering relation. The probability distribution of this "riskier" random variable has a simpler structure. Goovaerts et al. (1999) consider distributions of the present value of cash flows based on stochastic interest. They conclude that the comonotonic joint distribution (the distribution that is the largest in convex order) is often a good approximation of the original distribution. The latter can usually only be derived by means of simulation.

The aim of this paper is to investigate if something similar applies for a multistate life insurance contract. The quality of the approximation by means of the comonotonic joint distribution will be analyzed by comparing it with the present value distribution derived by means of the algorithm developed by Hesselager & Norberg (1996). The latter algorithm is executed in such a way that the present value distribution thus obtained is very close to the real one. For the sake of simplicity, we will restrict ourselves to contracts involving only benefits due in case of remaining in a state.

The paper is organized as follows. In Section 2, the basic assumptions are stated and the relevant present value random variable, the distribution of which is to be derived, is displayed. A summary of the algorithm developed in Hesselager & Norberg (1996) is presented in Section 3. In Section 4, an alternative way to evaluate the distribution of a present value random variable is discussed. Besides, for some disability annuities, we derive the comonotonic joint distribution of the payments and analyze its properties. Section 5 contains illustrating examples and Section 6 concludes.

2 Basic assumptions

The following assumptions are the same as in Hesselager & Norberg (1996), except that benefits and premiums due to transition from one state to another are not taken into account.

Consider a set \( \zeta = \{0, \ldots, J\} \) of all possible states of a general life policy, such that at any time \( t \in [0, n] \), the policy is in one and only one state. The state of the policy at time \( t \) is denoted by \( X(t) \). The stochastic process is taken to be right continuous with \( X(0) = 0 \), implying that the policy is in state 0 at time 0, being the time-upon-issue. Introduce \( I_j(t) \) as the indicator of the event that the contract is in state \( j \) at time \( t \), and

\[
N_{jk}(t) = \# \{ \tau \in (0, t) \mid X(\tau-) = j, X(\tau) = k \}
\]

(1)
as the total number of transitions of \( \{X(t)\}_{t \geq 0} \) from state \( j \) to state \( k \) (\( \neq j \)) by time \( t \). The payment function \( B \) is assumed to be continuous from the right as well. It is specified as

\[
dB(t) = \sum_j I_j(t-)dB_j(t)
\]

(2)

where each \( B_j \) is a deterministic payment function specifying payments due during sojourns in state \( j \) (a general life annuity). The left limit in \( I_j(t-) \) means that the state annuity is effective at time \( t \) if the policy is in state \( j \) just prior to (but not necessarily equal at) time \( t \). Consistent with this, we define \( I_0(0-) = 1 \). Specification (2) implies that there are no benefits due to transition from one state to another.

We assume that each \( B_j \) decomposes into an absolutely continuous part and a discrete part as

\[
dB_j(t) = b_j(t) \, dt + \Delta B_j(t)
\]

(3)
Premiums are counted as negative benefits.

It is supposed that \( \{X(t)\}_{t \geq 0} \) is a Markov chain. Denote the transition probabilities by
\[
p_{jk}(t, u) = P[X(u) = k | X(t) = j].
\]  \( \tag{4} \)

The transition intensities
\[
\mu_{jk}(t) = \lim_{h \to 0} \frac{p_{jk}(t + h, u) - p_{jk}(t, u)}{h}
\]
\( \tag{5} \)

are assumed to exist for all \( j, k \in \zeta, j \neq k \). The total intensity of transition out of state \( j \) is \( \mu_j(t) = \sum_{k \neq j} \mu_{jk}(t) \). The probability of staying uninterrupted in the state \( j \) during the time interval from \( t \) to \( u \) is
\[
e^{-\int_t^u \mu_j(s) \, ds}.
\]

We assume \( \delta(t) \), being the force of interest at time \( t \), to be a deterministic and continuous function. We introduce the following discount function:
\[
\nu(t) = e^{-\int_t^u \delta(s) \, ds}.
\]  \( \tag{6} \)

Our aim is to derive the distribution function of the random present value, which will be specified by \( V \). So
\[
V = \int_0^n \nu(t) \, dB(t).
\]  \( \tag{7} \)

Without loss of generality, single premiums (that are to be paid upon issue) will not be considered to be part of the present value.

3 The method by Hesselager & Norberg

The formulas displayed in this section are adopted from Hesselager & Norberg (1996). The difference in the formulas concerns the force of interest which in their paper is allowed to be state-dependent and fixed, conditionally given the state. In this contribution, on the other hand, the force of interest is taken deterministic and independent of the state where a contract remains, though not necessarily constant as a function of time. Recall, furthermore, that contracts with benefits due to transition are not taken into consideration. Hesselager & Norberg (1996) introduce the state-dependent probability functions
\[
P_j(t, u) = P \left[ \int_t^u \frac{\nu(r)}{\nu(t)} \, dB(t) \leq u | I_j(t) = 1 \right], t \in [0, n], u \in \mathbb{R}, j \in \zeta.
\]  \( \tag{8} \)

So \( P_j(t, u) \) denotes the probability that at time \( t \), conditionally given that the contract is then in state \( j \), the present value of future payments, discounted to \( t \), is smaller than or equal to \( u \).

Starting point in their analyses is the recursive equation
\[
P_j(t, u) = \sum_{k \neq j} \int_t^u e^{-\int_t^s \mu_{jk}(r) \, dr} \, ds \cdot P_k \left( \int_t^u \frac{\nu(r)}{\nu(s)} \, dB_j(t) \right)
\]
\[+ e^{-\int_t^u \mu_j(s) \, ds} I \left[ \int_t^u \frac{\nu(r)}{\nu(t)} \, dB_j(t) \leq u \right].
\]  \( \tag{9} \)
Applying the auxiliary function
\[
Q_j(t, u) = P_j \left( t, \nu(t)^{-1} \left( u - \int_0^t \nu(\tau) \, dB_j(\tau) \right) \right),
\] (10)
and substituting this in (9) yields
\[
e^{-\frac{1}{2} \mu(\nu) dt} Q_j(t, u) = \int e^{-\frac{1}{2} \mu(\nu) dt} \sum_{h,k \neq j} \mu_{jk}(s) \, ds \cdot Q_k \left( s, u + \int_0^t \nu(\tau) \, dB_k(\tau) - B_j(\tau) \right) + e^{-\frac{1}{2} \sigma(\nu) dt} I \left[ \int_0^t \nu(\tau) \, dB_j(\tau) \leq u \right].
\] (11)
By differentiating with respect to \( t \) and rearranging a bit it follows that the functions \( Q_j(t, u) \) in (11) are the unique solutions to the differential equations
\[
dQ_j(t, u) = \mu_j(t) \, dt \cdot Q_j(t, u) - \sum_{h,k \neq j} \mu_{jk}(t) \, dt \cdot Q_k \left( t, u + \int_0^t \nu(\tau) \, dB_k(\tau) - B_j(\tau) \right),
\] (12)
subject to the constraint
\[
Q_j(n, u) = I \left[ \int_0^n \nu(\tau) \, dB_j(\tau) \leq u \right].
\] (13)
The computational scheme follows by taking the finite difference version of the above equation:
\[
Q_j^*(t+h, u) = (1 - \mu_j(t) \cdot h) \cdot Q_j^*(t, u) + h \sum_{h,k \neq j} \mu_{jk}(t) \cdot Q_k^* \left( t, u + \int_0^t \nu(\tau) \, dB_k(\tau) - B_j(\tau) \right).
\] (14)
Then starting from (13) (with \( Q_j^* \) in the place of \( Q_j \)) one calculates first the functions \( Q_j^*(n - h, \cdot) \) by (14) and continues recursively until \( Q_j^*(0, \cdot) \) finally can be calculated. In the paper, the functions \( Q_j^*(t, u) \) are defined for \( t \in \{0, h, 2h, ..., n\} \) and \( u \in \{a, a+h', a+2h', ..., b\} \), where \( h \) and \( h' \) are certain step-lengths.
This method is the only general approach to tackle the problem of determining the present value distributions that has appeared in the literature until now. For practical calculations, however, it has the drawback that the method is time-consuming if only one needs to calculate the probability distribution at issue for a certain given state \( j \), that is, \( P_j(0, \cdot) \). If all the states in the contract are intercommunicating (this means that from one state, one can make a transition to each other state and vice versa) at any time the contract is valid, the total number of times that \( Q_j^*(t-h, u) \) in (14) needs to be computed is in the order of
\[
\frac{n}{h} \cdot \frac{b-a}{h'} \cdot M,
\] (15)
where $M$ denotes the total number of states. In (15), $\frac{a}{b}$ and $\frac{b-a}{b}$ represent the total values that the first and second argument of $Q^*$ can adopt, respectively. For instance, if $n = 30, \quad h = \frac{1}{100}, \quad b - a = 10, \quad h^* = \frac{1}{100}$ and $M = 3$, the relevant quantity needs to be calculated for about 180 million times. Besides, in each of these evaluations the second argument of the $Q^*$ function in the second term of the right hand side of (14) must be rounded to the nearest point in \{ $a, a + h, a + 2h, \ldots, b$ \}, thus slowing down the speed of computing. In practice, with at least one absorbing state (death) and usually some strongly transient states, fewer calculations have to be done, but the method can still be cumbersome.

In the next section, we will present a way to derive, for some disability annuity contracts, an approximation of the c.d.f. of $V$ defined in (7).

4 The probability distribution of a present value, largest in convex order

In this section we show that there are alternative ways to obtain the present value distribution of $V$ if the underlying Markov chain is relatively simple. First, in Subsection 4.1, we will illustrate this by means of an example, concerning a hierarchical Markov chain with only a few number of states. In Subsection 4.2, we will briefly discuss the approach in the paper Goovaerts et al. (1999) and rewrite the present value random variable in equation (7), such that their method can be applied.

4.1 Example of a contract under a hierarchical Markov chain

Consider a disability annuity equal to unity to be paid out continuously while the individual is disabled. The contract is issued at time 0, when the insured is active. Hence the number of states is equal to 3. Let the symbols $a$, $i$ and $d$ represent the three relevant states in this contract, namely "Active", "Disabled" and "Dead". The Markov chain is hierarchical if an individual cannot recover once being disabled. The situation can be displayed graphically as in Figure 1. So the $B_i(t)$ in the right hand side of (2) are equal to (recall that single premium payments are not taken into consideration)

$$ B_a(t) = B_d(t) = 0 \quad \forall t, \quad B_i(t) = \begin{cases} 0 & t < 0 \\ t & t \geq 0 \end{cases}. \quad (16) $$
Furthermore we assume $\delta(\cdot)$ to be constant and equal to $\delta$. Hence $\nu(t) = e^{-\delta t}$. For an active individual, the probability distribution of the present value at issue proves to be

$$P_a(0, u) = \begin{cases} 
eu \& \mu_a(\cdot) dr & 0 \\ \frac{\mu_a(\cdot)}{\mu_a(\cdot)} ds & 0 < u < 0; \\ \gamma(u) & u = 0; \\ 1 & 0 < u < H(n); \\ 0 & u \geq H(n). \end{cases} \tag{17}$$

with

$$\gamma(u) = \int_0^u \int_{\tau_1 = 0}^{\tau_{1}^{\max}(t)} e^{-\int_{\tau_1}^{\tau_{1}^{\max}(t)} \mu_a(\cdot) d\tau} \mu_a(\tau_1) e^{-\int_{\tau_1}^{\tau_2(t; \tau_1)} \mu_a(\cdot) d\tau} \mu_a(\tau_2(t; \tau_1)) d\tau_1 d\tau_2,$$

$$\tau_{1}^{\max}(t) = -\frac{\ln(e^{-\delta t} + \delta t)}{\delta}, \tau_2(t; \tau_1) = -\frac{\ln(1 + e^{-\delta t} - e^{-\delta \tau_1} + \delta t)}{\delta},$$

and $H(n) = \int_0^n e^{-\delta t} ds. \tag{18}$

We will give a short explanation:

1. The present value can never be negative;

2. $P_a(0, 0)$ is the probability that the contract will never enter the state "Disabled". This is the case if the insured either survives the entire period $[0, n]$ or dies as an active person;

3. $H(n)$ is the maximum possible present value, corresponding with the event that the individual gets disabled immediately after issue and is still disabled upon the end of the contract period;

4. If $0 < u < H(n)$, the individual, at some time the contract is in force, enters the state "Disabled" remains there for some time and enters the state "Dead" before the insurance treaty expires. The cumulative probability distribution is obtained by transforming the present value $u$ in the period the contract is in the state "Disabled", bordered by two points-of-time, namely the time-upon-entering the state $t$ and the time-upon-leaving that state, in the above specification denoted by $\tau_1$ and $\tau_2$, respectively.

So we see in the above example that there are cases where one does not need to be restricted to the method by Hesselager and Norberg. The derivation of the above c.d.f. is quite straightforward because the present value depends only on two variables: the time-of-getting-disabled and the time-of-death. This is due to the fact that:

1. there is only a few number of states applying, and that

2. the Markov chain is hierarchical: the state "Disabled", once left, cannot be visited again.

If, on the other hand, reactivation is possible, during the contract period an insured can be disabled for more than one time interval. This makes calculations such as above much more complicated.

However, even if a Markov chain is hierarchical, deriving the present value distribution is quite complicated if there are many states applying. An example is the multilife insurance contract, as given by Hesselager & Norberg (1996).
Our method of dealing with problems such as above involves replacing the random variable of present value by one being larger in convex order. If $X$ and $Y$ are random variables, $X$ precedes $Y$ in convex order (notation $X \leq_c Y$) if $E[f(X)] \leq E[f(Y)]$ for each convex function $f$. This method has already been applied for distributions of the present value of cash flows based on stochastic interest, see Goovaerts et al. (1999). A short overview of this paper is presented below.

4.2 Approach by Goovaerts et al. (1999)

Goovaerts et al. (1999) specify the present value random variable, again denoted by $V$, as a sum of r.v.'s, each of them involving a certain point of time:

$$V = \sum_{k=1}^{m} Y_k,$$

with

$$Y_k = \beta_k e^{-X_{t_k}}, \beta_k \in \mathbb{R}.$$

where $X_{t_k}$ represents the force of interest integrated from $0$ to $t_k$, so is actually the stochastic discounting factor of a payment at time $t_k$. Furthermore, $\beta_k$ is the payment itself. If one wants to obtain an expression for the probability distribution of $V$, one needs to realize that the $X_{t_k}$'s are not mutually independent. As a consequence, the variables $Y_k$ are not independent of one another either.

The authors first derive the distribution, which, within the class of random vectors $(Y_1, ..., Y_m)$ with fixed marginals (such a class is called a Fréchet class) is the comonotonic one. This distribution is the largest in convex order. If $F_1, ..., F_m$ are the c.d.f.'s of the respective r.v.'s $Y_1, ..., Y_m$, this comonotonic joint distribution of $Y_1, ..., Y_m$ is equal to the distribution of the random vector

$$(F_1^{-1}(U), ..., F_m^{-1}(U)),$$  \hspace{1cm} (21)

where $U \sim \text{Uniform}(0, 1)$ and $F_k^{-1}(u), k \in \{1, ..., m\}$, is defined by

$$F_k^{-1}(u) = \min \{x | F_k(x) \geq u\}.$$  \hspace{1cm} (22)

The r.v. which is the sum of the components in (21) is denoted by $W$, and:

$$V \leq_c W.$$  \hspace{1cm} (23)

Let $W_k = F_k^{-1}(U), k \in \{1, ..., m\}$, so $W = \sum_{k=1}^{m} W_k$. Then the joint c.d.f. of $W_1, ..., W_m$ is known to be

$$\Pr[W_1 \leq y_1, ..., W_m \leq y_m] = \min_{k \in \{1, ..., m\}} F_k(y_k).$$  \hspace{1cm} (24)

Let $F_W(\cdot)$ be the c.d.f. of $W$ and $F_W^{-1}(\cdot)$ its inverse, the latter defined in the same way as in (22). The c.d.f. of $W$ follows implicitly from the relationship

$$F_W^{-1}(u) = \sum_{k=1}^{m} F_k^{-1}(u), \quad u \in [0, 1].$$  \hspace{1cm} (25)

The quality of the approximation by means of this c.d.f can be analyzed by comparing it with the joint distribution obtained e.g. by means of Monte Carlo simulation.

The aim of our paper is to investigate if such a method also works well if one wants to derive the probability distribution of the present value of a multistate life insurance contract, as introduced
above. This will be done by replacing $V$ in (7) by a random variable which has the cocomonotonic joint distribution within the given Fréchet class. We will do this for two level disability annuities, one paid by single premium and the other paid by level premium while remaining in the state "Active". In the numerical examples in Section 6, which are based on these annuities, the quality of the approximation will be judged by comparing the c.d.f. of the latter random variable with an accurate approximation of the c.d.f. of $V$. The latter approximation is obtained by applying the algorithm of Hesselager and Norberg with small values for the step-lengths $h$ and $h'$, as defined in Section 3.

In the remainder of this paper it is assumed that the interval $[0,t_n]$ can be partitioned into subsequents intervals $[0,t_1],[t_1,t_2],...,[t_{m-1},t_m]$, such that there may only be benefits due to remaining in a certain state at points of time $t_1,t_2,...,t_{m-1},t_m=n$.

As a consequence, we can write $V$ in (7) as a sum of random variables:

$$V = \sum_{k=1}^{m} Y_k,$$

where

$$Y_k = \sum_{j \in \zeta} f_j(t_k) I_j(t_k),$$

with

$$f_j(t_k) = \nu(t_k) (B_j(t_k) - B_j(t_{k-1})), \quad j \in \zeta, \quad k \in \{1,...,m\}. \quad (28)$$

In (28), $f_j(t_k)$ denotes the present value of the benefit paid at time $t_k$ in case of remaining in state $j$.

**Remark 1** The fully continuous case arises as a special case if we let $m \to \infty$ and besides $\max_{k \in \{0,...,m-1\}} |t_{k+1} - t_k| \to 0$.

So $V$ can be decomposed in $m$ separate random variables where the $Y_k$ denotes the stochastic present value of benefits due to remaining in a certain state at time $t_k$, $k \in \{1,...,m\}$. We have

$$\Pr[V = z] = \sum_{Y_1=0}^{f_j(t_1)} \Pr[Y_1 = f_j(t_1), Y_2 = f_j(t_2), ... , Y_m = f_j(t_m)]$$

$$= \sum_{j \in \zeta} f_j(t_k) I_j(t_k) \cdot \Pr[X(1) = j, X(2) = j, ..., X(m) = j]. \quad (29)$$

We define $F_k$ as the c.d.f. of $Y_k$, $k \in \{1,...,m\}$. Furthermore, we define, just as above, $W_k = F_k^{-1}(U), k \in \{1,...,m\}$, with $U \sim \text{Uniform}(0,1)$. So $W = \sum_{k=1}^{m} W_k$ and

$$F_k(y_k) = \Pr[Y_k \leq y_k] = \Pr[W_k \leq y_k], \quad k \in \{1,...,m\}. \quad (30)$$

Note that these marginals have a support consisting of a finite number of points.

In the next section we will derive the cocomonotonic joint distribution for two basic disability annuities by applying formula (24). Besides, we will analyze the properties of this cocomonotonic joint distribution, under additional conditions regarding the several probabilities to be disabled at the respective points of time. In all the cases and examples following, it is assumed that recovery from disability is possible.
5 Two disability annuities

In this section we will apply the theory considered just before to two disability annuities. We suppose that the same three states a ("Active"), i ("Disabled") and d ("Dead"), as considered and defined in the case in Subsection 4.1, apply and that upon issue, the contract is in the state "Active". The basic difference with the given case is that it is assumed that recovery from disability is possible. Hence we are dealing with a Markov chain as displayed graphically in Figure 2.

Furthermore, in both cases there are no payments due if the individual is dead. So we have:

\[ f_a(t_k) = B_d(t_k) = 0; \quad f_i(t_k) = \nu(t_k) (B_i(t_k) - B_i(t_{k-1})) \geq 0, \quad \forall k \in \{1, ..., m\}. \quad (31) \]

The only difference between the two policies is the method of premium payment. In Subsection 5.1 we will deal with payment by single premium, while Subsection 5.2 will consider the case where premiums are paid on as long as the individual is active. In both subsections we will treat the fully continuous annuity (where all benefits are paid on a continuous base) as a special case.

We will conclude this section with Subsection 5.3, containing some remarks concerning the calculation of the transition probabilities \( p_{au}(0,\cdot) \), \( p_{ai}(0,\cdot) \) and \( p_{ad}(0,\cdot) \).

5.1 Payment by single premium

Recall that, by assumption, present values do not consist of single premium payments, so:

\[ B_a(t_k) = f_a(t_k) = 0 \quad \forall k \in \{1, ..., m\}. \quad (32) \]

In the given case the marginals \( F_k(y_k) \) in formula (24) are specified as:

\[ F_k(y_k) = \begin{cases} 
0 & \text{for } y_k < 0 \\
1 - p_{au}(0,t_k) & \text{for } 0 \leq y_k < f_i(t_k), \quad k \in \{1, ..., m\}. \\
1 & \text{for } y_k \geq f_i(t_k). 
\end{cases} \quad (33) \]

In practice, disability annuities are often contracts valid for the period that an individual is not retired. As a consequence, for \( n \) not too large (otherwise the death rates will dominate), \( p_{au}(0,t) \) is usually an increasing function of \( t \) for \( t \in (0,n) \). This yields the following theorem:

**Theorem 2** If \( \frac{d p_{au}(0,t)}{dt} \geq 0 \), the comonotonic joint distribution of the present value corresponding to a disability annuity contract, paid against single premium, with the payment scheme as
specified in (31) and (32) is:

$$
\Pr [W \leq w] = \begin{cases} 
0 & \text{for } w < 0; \\
1 - p_{in} (0, t) & \text{for } \sum_{k=1}^{m} f_{t} (t_{k}) \leq w < \sum_{k=1}^{m} f_{t} (t_{k}) , \quad \ell \in \{1, \ldots, m\} ; \\
1 & \text{for } w \geq \sum_{k=1}^{m} f_{t} (t_{k}) .
\end{cases}
$$

(34)

Proof. Note that, for any \( y_{\ell} < 0 \) with \( \ell \in \{1, \ldots, m\} \):

$$
F_{W_{1}, \ldots, W_{m}} (y_{1}, \ldots, y_{m}) = F_{W_{1}, \ldots, W_{m}} (0, \ldots, 0, y_{\ell}, 0, \ldots, 0) = 0 .
$$

(35)

Besides, if \( p_{in} (0, t) \) is increasing in \( t \in [0, n] \) we have for \( y_{1}, \ldots, y_{\ell-1} \geq 0 \) and \( 0 \leq y_{\ell} < f_{t} (t_{\ell}) , \quad \ell \in \{2, \ldots, m\} \), that

$$
F_{W_{1}, \ldots, W_{m}} (y_{1}, \ldots, y_{m}) = F_{W_{1}, \ldots, W_{m}} (0, \ldots, 0, y_{\ell+1}, \ldots, y_{m}) .
$$

(36)

The above equality states that if an individual is not disabled at time \( t_{\ell} \), he cannot be disabled before that time either. This is equivalent to saying that if an individual is disabled at time \( t_{\ell} \), he will remain so with certainty till the expiration of the contract. This implies that the present value only depends on the time at which the contract enters the state "Disabled". This proves the theorem. \( \blacksquare \)

Note that the Markov chain corresponding to the distribution of \( W \) is hierarchical. The chain is even more rigorous than the ordinary hierarchical Markov chain considered in Subsection 4.1: a disabled individual can neither recover nor die. In other words: compared to the ordinary hierarchical Markov chain, the state "Disabled" is an absorbing one and not a strongly transient one.

On the other hand, since the marginals are fixed, the probabilities to get disabled are lower and the death rates for an active person are higher.

The fully continuous version of (34) leading to a benefit payment of \( dB_{t} (t) \) if the contract is in state \( t \) at time \( t \), with \( t \in (0, n] \) (so \( dB_{t} (t) > 0 \) on the same interval \( t \in (0, n] \)) is obtained by letting \( m \rightarrow \infty \) and besides \( \max_{t \in [0, m-1]} [t_{t+1} - t_{t}] \rightarrow 0 \) (cf. Remark 1). The result is:

$$
\Pr [W \leq w] = \begin{cases} 
0 & \text{for } w < 0; \\
1 - p_{in} (0, t) & \text{for } w = \int_{0}^{t} f_{t} (s) \, ds , \quad t \in (0, n] ; \\
1 & \text{for } w > \int_{0}^{t} f_{t} (s) \, ds .
\end{cases}
$$

(37)

One of the numerical examples in Section 6 will be based on this contract.

The next equations show to illustrate the transition intensities corresponding to the commonotonic joint distribution. These are obtained from the forward differential equations of Chapman-Kolmogorov. For notational convenience, they are accompanied by an asterisk superscript (*). The original transition intensities, corresponding to the joint distribution of \( V \), are given between
\[ \mu_{ad}(t) = 0 \]
\[ \mu_{id}(t) = 0 \]
\[ \mu_{aa}(t) = \frac{1}{p_{aa}(0,t)} \frac{d p_{aa}(0,t)}{dt} + p_{aa}(0,t) (\mu_{ad}(t) + \mu_{dd}(t)) \]
\[ \mu_{ai}(t) = \frac{1}{p_{aa}(0,t)} \frac{d p_{ad}(0,t)}{dt} \]
\[ \mu_{ia}(t) = \frac{1}{p_{aa}(0,t)} \frac{d p_{ai}(0,t)}{dt} \]
\[ \mu_{ad}(t) = \frac{1}{p_{aa}(0,t)} \frac{d p_{id}(0,t)}{dt} + p_{aa}(0,t) (\mu_{ad}(t) + \mu_{dd}(t)) \]

Next, we will consider annuity treaties where premiums are paid while the contract is in the state "Active". We will treat this topic in the same way as above.

### 5.2 Premium payment in state "Active"

In this case the premiums discounted to time-upon-issue are:

\[ f_a(t_k) = \nu(t_k)(B_a(t_k) - B_a(t_{k-1})) \leq 0, \quad \forall k \in \{1, \ldots, m\}. \]  

The marginals \( F_k(y_k) \) in equation (24) are, as one might expect, a bit more complicated than in the single premium case:

\[ F_k(y_k) = \begin{cases} 0 & \text{for } y_k < f_a(t_k); \\ 1 - p_{aa}(0,t_k) - p_{ad}(0,t_k) & \text{for } f_a(t_k) \leq y_k < 0; \\ 1 - p_{aa}(0,t_k) & \text{for } 0 \leq y_k < f_i(t_k); \\ 1 & \text{for } y_k \geq f_i(t_k). \end{cases} \]

We again assume that \( p_{aa}(0,t) \) is an increasing function of \( t \) for \( t \in (0, n) \), leading to the following theorem:

**Theorem 3** If \( \frac{d p_{aa}(0,t)}{dt} \geq 0 \), the comonotonic joint distribution of the present value corresponding to a disability annuity contract, with the payment scheme as specified in (31) and (42) is:

\[ \Pr[W \leq w] = \begin{cases} 0 & \text{for } w < \sum_{k=1}^{m} f_a(t_k); \\ 1 - p_{aa}(0,t_\ell) - p_{ad}(0,t_\ell) & \text{for } w \in [g_a(\ell), g_a(\ell - 1), r \in \{1, \ldots, m\}, r \in \{r_{\min}, \ldots, m\}; \\ 1 & \text{for } w \geq \sum_{k=1}^{m} f_i(t_k). \end{cases} \]
In the above formulas
\[
       g_{\ell}(r) = \sum_{k=1}^{r} f_{\ell}(t_k); \\
       g_{\ell}(r) = \sum_{k=r+1}^{m} f_{\ell}(t_k) + \sum_{k=1}^{r} f_{\ell}(t_k); \\
       r_{\min} = \max \left[ r \in \{1, ..., m-1\} | \sum_{k=r+1}^{m} f_{\ell}(t_k) + \sum_{k=1}^{r} f_{\ell}(t_k) < 0 \right]. \tag{45}
\]

Proof. Note that, for any \( y_t < f_{\ell}(t_t) \) with \( \ell \in \{1, ..., m\} \):
\[
   F_{W_1, ..., W_m}(y_1, ..., y_m) = F_{W_1, ..., W_m}(0, ..., 0, y_t, 0, ..., 0) = 0. \tag{46}
\]
Furthermore, if \( p_{\ell_0}(0, t) \) is increasing in \( t \in [0, n] \) we have the following:

1. If \( y_j \geq f_{\ell}(t_t) \) and \( f_{\ell}(t) \leq y_t < 0 \) for \( j \in \{1, ..., \ell - 1\} \) and \( \ell \in \{2, ..., m\} \) then
\[
   F_{W_1, ..., W_m}(y_1, ..., y_{\ell-1}, y_\ell, ..., y_m) = F_{W_1, ..., W_m}(f_{\ell}(t_t), ..., f_{\ell}(t) , y_{\ell+1}, ..., y_m), \tag{47}
\]
   implying that if an individual is active at a certain time, he is active all the time before. The consequence is that a disabled individual cannot recover.

2. For \( y_1, ..., y_{\ell-1} \geq 0 \) and \( 0 \leq y_t < f_{\ell}(t_t) \), \( \ell \in \{2, ..., m\} \), that
\[
   F_{W_1, ..., W_m}(y_1, ..., y_m) = F_{W_1, ..., W_m}(0, ..., 0, y_t, 1, ..., y_m). \tag{48}
\]
The above equality states that if an individual is not disabled at time \( t_t \), he cannot be disabled before that time either. This is equivalent to saying that if an individual is disabled at time \( t_t \), he will remain so with certainty till the expiration of the contract. This implies that the present value only depends on the time at which the contract enters the state "Disabled" or the time at which the contract enters the state "Dead".

This proves the theorem.  \( \blacksquare \)

The fully continuous version of (44) leading to a benefit payment of \( dB_t(t) \) if the contract is in state \( t \) and a premium payment of \(-dB_a(t) \) if the contract is in state \( a \) at time \( t \), with \( t \in (0, n) \) (so \( dB_t(t) > 0 \) and \( dB_a(t) < 0 \) on the same interval \( t \in (0, n) \), is, just as in the case of single premium payment, obtained by letting \( m \to \infty \) and besides \( \max_{a \in \{0, ..., m-1\}} \{t_{a+1} - t_t\} \to 0 \) (cf. Remark 1). The result is:
\[
   \Pr[W \leq w] = \begin{cases} 
   1 & \text{for } w < \int_0^t f_{\ell}(s) \, ds; \\
   1 - p_{\ell_0}(0, g_{\ell}(w)) - \int_0^t f_{\ell}(s) \, ds \leq w < 0; \\
   1 - p_{\ell_0}(0, g_{\ell}(w)) & \text{for } 0 \leq w < \int_0^t f_{\ell}(s) \, ds; \\
   1 & \text{for } w \geq \int_0^t f_{\ell}(s) \, ds. 
\end{cases} \tag{49}
\]
In the above formula \( g_{\ell}(u) \) and \( g_{\ell}(u) \) are the solutions of \( t \) in the equalities
\[
   \int_t^u f_{\ell}(s) \, ds + \int_0^t f_{\ell}(s) \, ds = w \tag{50}
\]
and
\[
   \int_0^u f_{\ell}(s) \, ds = w, \tag{51}
\]
respectively.

One of the numerical examples in Section 6 will deal with the continuous case, and will therefore be based on equation (49).
5.3 Calculation of transition probabilities

Note that up to now we assumed that the transition probabilities were known. In fact, they need to be computed first by solving a system of forward differential equations of Chapman-Kolmogorov. It turns out that the method, in order to be computationally efficient, requires that the numerical calculation of the transition probabilities does not take too much time. With the availability of modern software packages (such as Mathematica) and powerful computers, this is likely to be the case, except when the number of states is very high. One can also assume that only a limited number of transitions can take place in a certain period of time. For instance, both in the AIDS model and in the disability annuity model used in the Netherlands (cf. Alting von Geusau, 1990, and Gregoiruis, 1993, respectively) it is supposed that there can be at most one transition per year.

In the next section, a numerical example will be displayed, illustrating the two theorems we just derived.

6 Numerical examples

An insurance contract pays an amount of 1 on a continuous basis while the insured is in the state disabled. Hence $B_{*}(t) = t; \ t \in [0, n]$. We use the same transition intensities as those applied in the numerical example 5.3 of Hesselager & Norberg (1996):

$$
\begin{align*}
\mu_{td}(t) &= \mu_{id}(t) = 0.0005 + 10^{-4.12 + 0.04(t(n+1))}; \\
\mu_{nu}(t) &= 0.005; \\
\mu_{ua}(t) &= 0.0004 + 10^{-5.46 + 0.06(t(n+4))}; \\
\eta &= \frac{1}{1000}; \ n = 30; \ \delta = \ln(1.045).
\end{align*}
$$

The value of $\delta$ corresponds to an annual level of interest of 4.5%. The specification of the transition intensities results in the transition probabilities $p_{ad}(0, t)$, $p_{ua}(0, t)$ and $p_{ud}(0, t)$ as displayed graphically in Figure 3. We have studied two separate ways of premium payment:

1. Payment by single premium;
2. Level premium payment such that there is equivalence at issue.

These two cases will be treated below.

Case 4 (Payment by single premium) Recall that we have assumed that single premiums are not part of the present value. Hence in case of single premium payment the minimal value of that present value is equal to 0 (corresponding to the event that the contract will never enter the state "Disabled") while the maximal value is equal to

$$
\int_{0}^{\infty} (1.045)^{t} \ dt = 16.6527.
$$

The algorithm by Hesselager and Norberg has been applied for $a = -0.01$, $b = 16.66$ and $h_0 = \frac{1}{7000}$, corresponding to 2281 points of support.

The distribution of $W$, displayed in formula (37), is as below:

$$
\Pr[W \leq w] = \begin{cases} 
0 & \text{for } w < 0; \\
1 - p_{ad}(0, t) & \text{for } w = \ln(1.045) \left( (1.045)^{t} - (1.045)^{w} \right) \ t \in [0, n]; \\
1 & \text{for } w > 16.6527.
\end{cases}
$$

Both the approximate c.d.f. of $V$ and the comonotonic c.d.f., exhibited in (24), are as displayed in Figure 4.
Figure 3: Graphical display of $p_{ad}(0,t)$ (dotted), $p_{al}(0,t)$ (solid) and $p_{ad}(0,t)$ (dashed) as a function of $t$.

Figure 4: C.d.f. of present value in case of single premium payment: approximate version of $V$, obtained by Hesselager & Norberg algorithm (solid curve) and comonetonic version $W$ (dotted curve).
Case 5 (Level premium payment)  In case of level premium payment on a continuous basis, with premium volume $c$, the premium payment function has the shape $B_t(t) = -ct; t \in [0,n]$. The level premium satisfying the principle of equivalence is in this case equal to

$$c = \frac{\int_0^n e^{-kt} p_{pt}(0,t) \, dt}{\int_0^n e^{-kt} p_{pt}(0,t) \, dt} = 0.0175456. \quad (55)$$

The maximum possible present value is the same as in the previous case, while the minimum, attained in case the individual remains in the state "Active" from the beginning till the end of the contract period, proves to be

$$-c \int_0^t e^{-kt} \, dt = a = -0.2922. \quad (56)$$

The algorithm by Hesselager and Norberg has in this case been used for $a = -0.35, b = 16.7$ and, just as in the previous case $b' = \frac{1000}{2000}$ corresponding to 2435 points of support. The distribution of $W$, displayed in formula (49), reads as:

$$\Pr[W \leq w] = \begin{cases} 1 - p_{pt}(0,g_t(w)) - p_{pt}(0,g_d(w)) & \text{for } w <-0.29; \\ 1 - p_{pt}(0,g_t(w)) & \text{for } -0.2922 \leq w < 0; \\ 1 & \text{for } w \geq 16.6527. \end{cases} \quad (57)$$

In (57),

$$g_t(w) = -\frac{\ln \left( \frac{e^{wt} + e^{-nt - w}}{1+c} \right)}{\delta}; g_d(w) = -\frac{\ln (1 + \frac{w}{\delta})}{\delta}. \quad (58)$$

Both the approximate c.d.f. of $V$ and the comonotonic c.d.f. exhibited in (57) are as given in Figure 5. For graphical reasons the plotting has been restricted to the region where the deviations between the two curves are relatively large.

One can see that, in both cases, and especially in the case of level premium payment, the approximation of the c.d.f. of $V$ by means of the c.d.f. of $W$ gives very good results. Besides, the approximation method has shown to be computationally very efficient. In both the cases of single premium and level premium payment, it took only a few minutes to derive the c.d.f. of $W$ while the algorithm by Hesselager and Norberg required many hours of calculation time.

7 Conclusions and future research

In this paper, we have shown that the method implemented by Goovaerts et al. (1999) to approximate the probability distribution of a present value by means of the c.d.f. of the random variable being the upper bound in convex sense can also be applied to some life insurance contracts with more than two states applying. In the cases considered the c.d.f. of the comonotonic joint distribution turns out to have a simple structure.

This paper is only a start-up for an extensive line of research. Extension in several ways is necessary. It has to be investigated what the consequences would be if more states were added to our model. Furthermore, the paper has dealt with insurance treaties where there are only benefits due in case of remaining in a state. Contracts involving benefits due in case of transition from one state to another (lump sum benefits e.g. due upon death of the insured) have not been considered. It would be interesting to find out what the consequences would be in case such lump sum benefits were also part of the insurance policy.

Finally, it would be nice if there were a general method to derive the c.d.f. being lowest in convex order. This would improve the testing of the approximation's quality.
Figure 5: C.d.f. of present value in case of level premium payment: approximate version of $V$, obtained by Hesselager & Norberg algorithm (solid curve) and comonotonic version $W$ (dotted curve).

Acknowledgment
The author thanks Angela van Heerwaarden, Rob Kaas and Henk Wolthuis for giving helpful comments.

References


DEPARTMENT OF ACTUARIAL SCIENCE AND STATISTICS

Actuarial Research Papers since 1995


ISBN 1 874 770 85 9
ISBN 1 874 770 88 3
ISBN 1 874 770 89 1
ISBN 1 874 770 90 5
ISBN 1 874 770 91 3
ISBN 1 874 770 93 X
ISBN 1 874 770 94 8
ISBN 1 874 770 99 9
ISBN 1 874 770 00 X
ISBN 1 901615 03 0
ISBN 1 901615 04 9
ISBN 1 901615 05 7
ISBN 1 901615 06 5
ISBN 1 901615 07 3
ISBN 1 901615 08 1
ISBN 1 901615 09 X
ISBN 1 901615 12 X


103. Rickayzen B.D. A Sensitivity Analysis of the Parameters used in a PHI Multiple State Model. December 1997. 18 Pages. ISBN 1 901615 20 0


   ISBN 1 901615 36 7

   ISBN 1 901615 38 3

   ISBN 1 901615 39 1

   ISBN 1 901615 41 3

   ISBN 1 901615 42 1

   ISBN 1 901615 43 X

   ISBN 1 901615 44 8

   ISBN 1 901615 45 6

   ISBN 1 901615 45 4

   ISBN 1 901615 47 2

   ISBN 1 901615 49 9

   ISBN 1 901615 50 2

Statistical Research Papers

   ISBN 1 874 770 83 2

   ISBN 1 874 770 86 7

   ISBN 1 874 770 87 5
<table>
<thead>
<tr>
<th></th>
<th>Title</th>
<th>Author(s)</th>
<th>Publication Date</th>
<th>Pages</th>
<th>ISBN</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>A Note on D-optimal Designs for a Logistic Regression Model</td>
<td>Sebastiani P. and Settimi R.</td>
<td>May 1996</td>
<td>12</td>
<td>1 874 770 92 1</td>
</tr>
<tr>
<td>5</td>
<td>First-order Optimal Designs for Non Linear Models</td>
<td>Sebastiani P. and Settimi R.</td>
<td>August 1996</td>
<td>28</td>
<td>1 874 770 95 6</td>
</tr>
<tr>
<td></td>
<td>and Control</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>Moments and Generating Functions for the Absorption Distribution</td>
<td>Newby M.</td>
<td>September 1996</td>
<td>16</td>
<td>1 874 770 97 2</td>
</tr>
<tr>
<td></td>
<td>and its Negative Binomial Analogue</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>Mixture Reduction via Predictive Scores</td>
<td>Cowell R.G.</td>
<td>November 1996</td>
<td>17</td>
<td>1 874 770 98 0</td>
</tr>
<tr>
<td>9</td>
<td>Robust Parameter Learning in Bayesian Networks with Missing Data</td>
<td>Sebastiani P. and Ramoni M.</td>
<td>March 1997</td>
<td>9</td>
<td>1 901615 00 6</td>
</tr>
<tr>
<td>10</td>
<td>Guidelines for Corrective Replacement Based on Low Stochastic</td>
<td>Newby M.J. and Coolen F.P.A.</td>
<td>March 1997</td>
<td>9</td>
<td>1 901615 01 4</td>
</tr>
<tr>
<td></td>
<td>Structure Assumptions</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>Approximations for the Absorption Distribution and its Negative</td>
<td>Newby M.J.</td>
<td>March 1997</td>
<td>6</td>
<td>1 901615 02 2</td>
</tr>
<tr>
<td></td>
<td>Binomial Analogue</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>The Use of Exogenous Knowledge to Learn Bayesian Networks from</td>
<td>Ramoni M. and Sebastiani P.</td>
<td>June 1997</td>
<td>11</td>
<td>1 901615 10 3</td>
</tr>
<tr>
<td></td>
<td>Incomplete Databases</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>Learning Bayesian Networks from Incomplete Databases</td>
<td>Ramoni M. and Sebastiani P.</td>
<td>June 1997</td>
<td>14</td>
<td>1 901615 11 1</td>
</tr>
<tr>
<td>14</td>
<td>Risk Based Optimal Designs</td>
<td>Sebastiani P. and Wynn H.P.</td>
<td>June 1997</td>
<td>10</td>
<td>1 901615 13 8</td>
</tr>
<tr>
<td>15</td>
<td>Sampling without Replacement in Junction Trees</td>
<td>Cowell R.</td>
<td>June 1997</td>
<td>10</td>
<td>1 901615 14 6</td>
</tr>
<tr>
<td>16</td>
<td>Optimal Overhaul Intervals with Imperfect Inspection and Repair</td>
<td>Dugg R.A. and Newby M.J.</td>
<td>July 1997</td>
<td>11</td>
<td>1 901615 15 4</td>
</tr>
<tr>
<td>17</td>
<td>Bayesian Experimental Design and Shannon Information</td>
<td>Sebastiani P. and Wynn H.P.</td>
<td>October 1997</td>
<td>11</td>
<td>1 901615 17 0</td>
</tr>
<tr>
<td>18</td>
<td>A Characterisation of Phase Type Distributions</td>
<td>Wolstenholme L.C.</td>
<td>November 1997</td>
<td>11</td>
<td>1 901615 18 9</td>
</tr>
<tr>
<td>19</td>
<td>A Comparison of Models for Probability of Detection (POD) Curves</td>
<td>Wolstenholme L.C.</td>
<td>December 1997</td>
<td>23</td>
<td>1 901615 21 9</td>
</tr>
<tr>
<td>20</td>
<td>Parameter Learning from Incomplete Data Using Maximum Entropy I:</td>
<td>Cowell R.G.</td>
<td>February 1999</td>
<td>19</td>
<td>1 901615 37 5</td>
</tr>
<tr>
<td></td>
<td>Principles</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>Parameter Learning from Incomplete Data Using Maximum Entropy II:</td>
<td>Cowell R.G.</td>
<td>November 1999</td>
<td>12</td>
<td>1 901615 40 5</td>
</tr>
<tr>
<td></td>
<td>Application to Bayesian Networks</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Department of Actuarial Science and Statistics

Actuarial Research Club

The support of the corporate members

Computer Sciences Corporation
Government Actuary’s Department
Guardian Insurance
Hymans Robertson
KPMG
Munich Reinsurance
PricewaterhouseCoopers
Swiss Reinsurance
Watson Wyatt

is gratefully acknowledged.

ISBN 1 901615 50 2