

STICHTING
MATHEMATISCH CENTRUM
2e BOERHAAVESTRAAT 49
AMSTERDAM
AFDELING MATHEMATISCHE STATISTIEK

S 367 (SP 106)
second version

Driving with Markov-programming

by

Prof.dr. G. de Leve

and

ir. P.J. Weeda



july 1967

The Mathematical Centre at Amsterdam, founded the 11th of February, 1946, is a non-profit institution aiming at the promotion of pure mathematics and its applications, and is sponsored by the Netherlands Government through the Netherlands Organization for the Advancement of Pure Research (Z.W.O.) and the Central Organization for Applied Scientific Research in the Netherlands (T.N.O.), by the Municipality of Amsterdam and by several industries.

1. The motorist problem

1.1 Introduction

Suppose we consider the following problem.

A motorist has decided to effect an accident insurance under the following conditions. The insurance runs for one year. The premium for the first year amounts E_0 . If no damages have been claimed during i successive years, $i = 1, 2$ or 3 the premium is reduced to E_i . After four years of damagefree driving no further premium reduction is granted, so the premium remains E_3 . The premium is due on the first day of the year ^{*)}. The own risk amounts a_0 .

The number of accidents of our motorist during a time period T is assumed to be Poisson distributed with parameter λT . The extend of the damage s ^{**)} has distribution function $F(s)$ with finite mean and variance.

The problem of our motorist will be to decide whether to claim a damage or not. He will have to develop a strategy that specifies his decisions in every possible situation. His strategy will be called optimal if it minimizes the expected costs in the long run.

We may expect that in view of the premium reduction, it will be unprofitable to claim damages which are not much larger than a_0 . Once a damage is claimed it will be profitable to claim all following damages that exceed a_0 during the remaining part of the year.

Hence his decisions will also depend on the time of the year and the premium paid at the beginning of that year. So we distinguish between four types of years, for each premium one.

Our task will be to determine for each type of year a function $s(t)$ with the following property: If at time t an accident occurs with damage s

*) It is no restriction to assume that this is January 1st.

***) Random variables are underlined.

and no damages have been claimed since the last payment of premium, then s should be claimed if $s > s(t)$.

The strategy is completely fixed by this function.

The optimal strategy will be the function $s(t)$ that minimizes the expected costs of the motorist.

In this paper, a mathematical method, called Markov-programming, is developed that yields the function $s(t)$ under the assumptions stated above. Before presenting a review of the method in part 2 and its application to the motorists problem in part 3, some numerical results will be given first.

1.2 Numerical results

Suppose the following numerical data are given

$$E_0 = 1.6$$

$$E_1 = 1.4$$

$$E_2 = 1.2$$

$$E_3 = 1.1$$

$$a_0 = 0.4$$

Primarily the effect of the damage distribution $F(s)$ on the optimal strategy is investigated. Three different types of distributions have been used:

a) the exponential distribution with density function

$$f(s) = \mu e^{-\mu s}$$

and expectation: $E \underline{s} = \frac{1}{\mu}$

variance: $\text{Var } \underline{s} = \frac{1}{\mu^2}$

coefficient of variation: $\alpha = \sqrt{\frac{\text{Var } \underline{s}}{(E \underline{s})^2}} = 1$

b) the gamma distribution with density function

$$f(s) = \frac{\mu^k s^{k-1}}{(k-1)!} e^{-\mu s}$$

(only integer values of k were considered here)

and expectation: $E \underline{s} = \frac{k}{\mu}$

variance: $\text{Var } \underline{s} = \frac{k}{\mu^2}$

coefficient of variation: $\alpha = \sqrt{\frac{\text{Var } \underline{s}}{(E \underline{s})^2}} = \frac{1}{\sqrt{k}}$

c) The lognormal distribution with density function

$$f(s) = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{(\ln s - \mu)^2}{2\sigma^2}} \cdot \frac{1}{s}$$

and expectation: $E \underline{s} = e^{\mu + \frac{1}{2}\sigma^2}$

variance: $\text{Var } \underline{s} = e^{2\mu + 2\sigma^2} - e^{2\mu + \sigma^2}$

coefficient of variation: $\alpha = \sqrt{\frac{\text{Var } \underline{s}}{(E \underline{s})^2}} = \sqrt{e^{\sigma^2} - 1}$.

We compare five distributions with the same expectation. Their density functions are sketched in figure 1.2.1. They are identified as follows:

Number of curve	Type of distribution	expectation	coefficient of variation
1	exponential	1	1
2	gamma	1	1/3
3	lognormal	1	1
4	"	1	1/3
5	"	1	3

The number of accidents in a year is Poisson distributed with expectation $\lambda = 2$. The corresponding optimal strategies are presented in figure 1.2.2. From these results it can be deduced that at least for

the distributions considered with the same mean and variance the optimal strategies are nearly the same. Further, increasing the variance leads to less conservative claiming in these cases.

Secondly the effect of the expected number of accidents is investigated. In figure 1.2.3 the optimal strategy is given for $\lambda = 2$ (curve 1) and $\lambda = 4$ (curve 6). In both cases the damage distribution is exponential with expectation 1. An increase of the number of accidents leads to more conservative claiming especially in the beginning of the year.

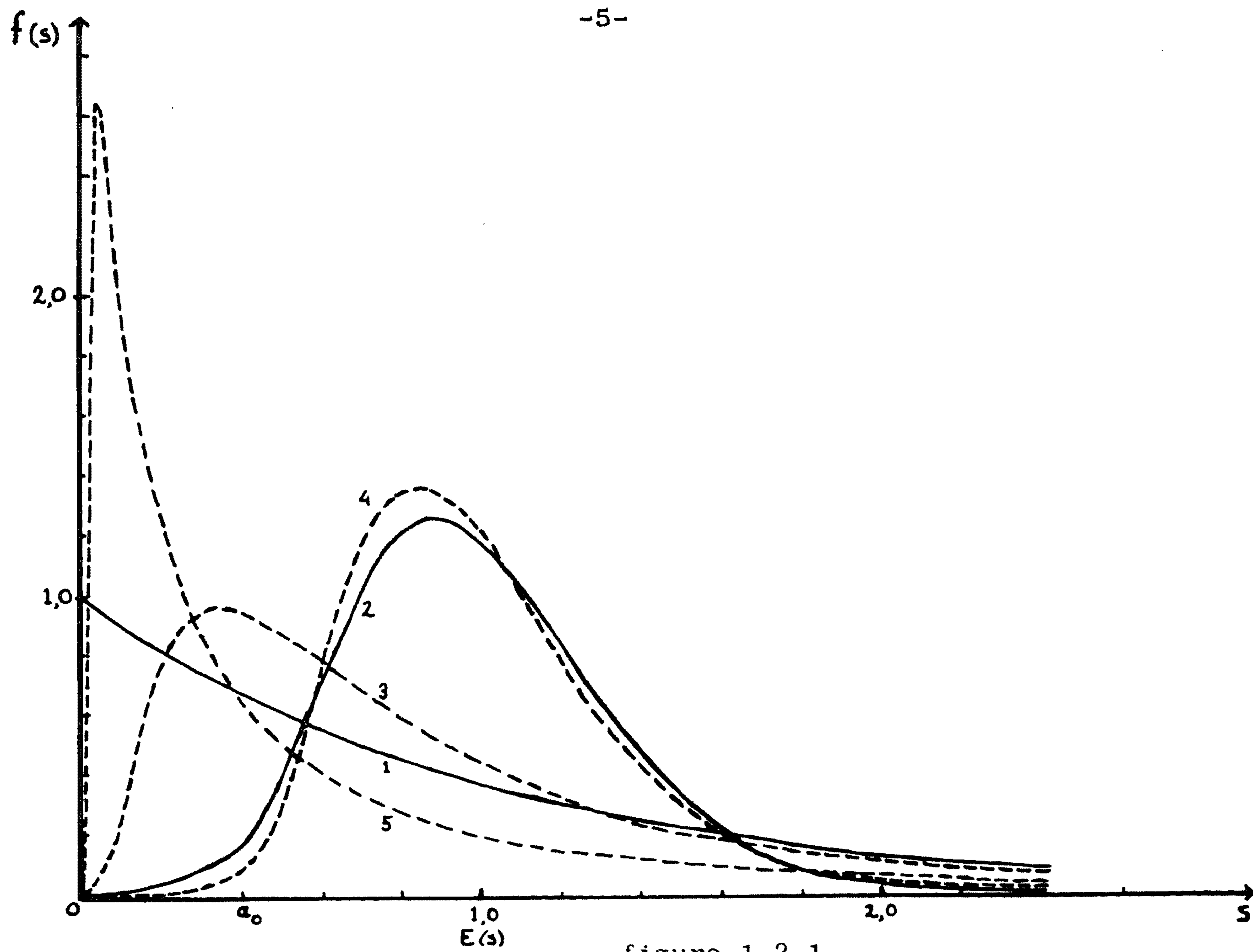


figure 1.2.1
The five used damage distributions

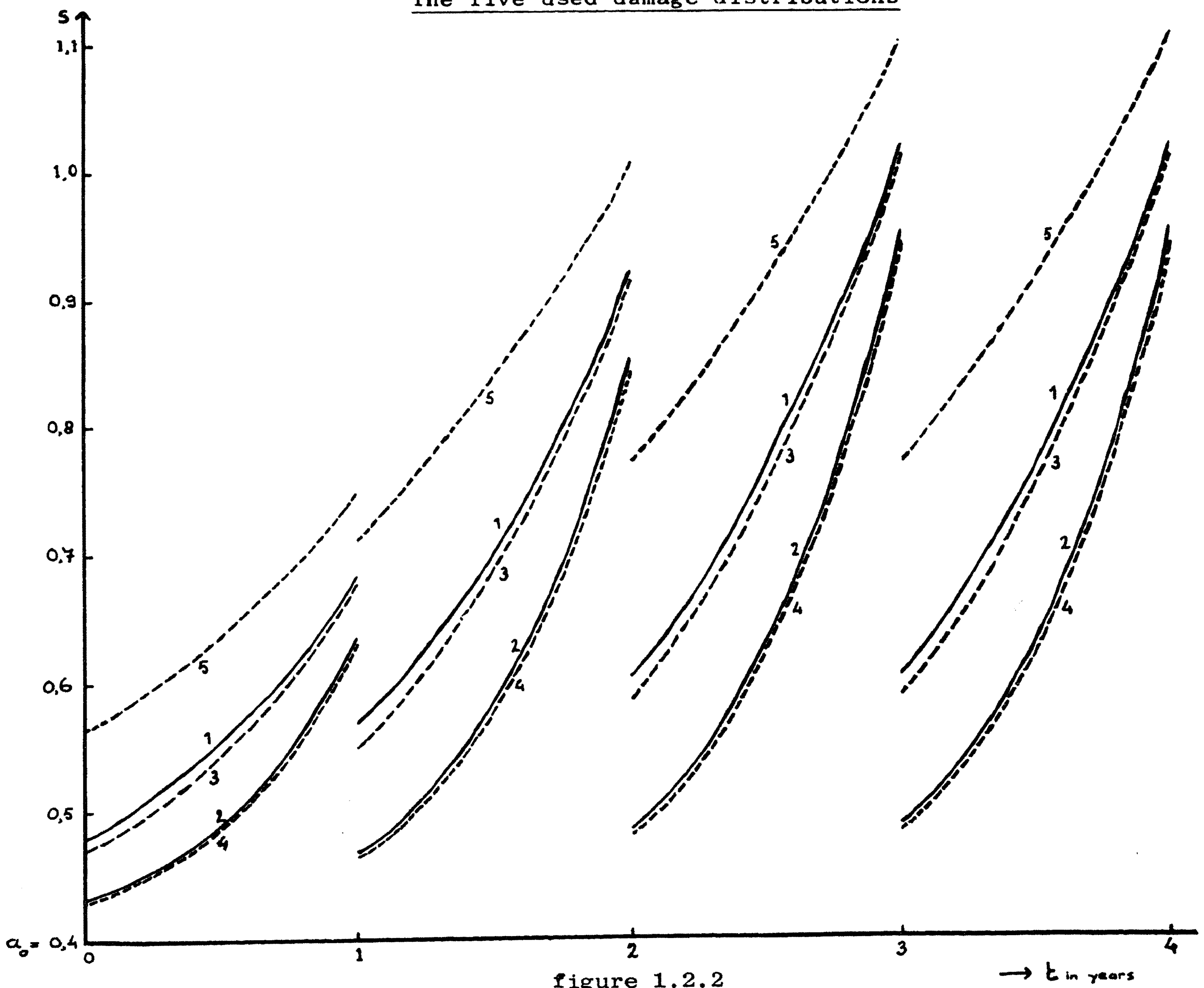


figure 1.2.2
The optimal strategy for these distributions

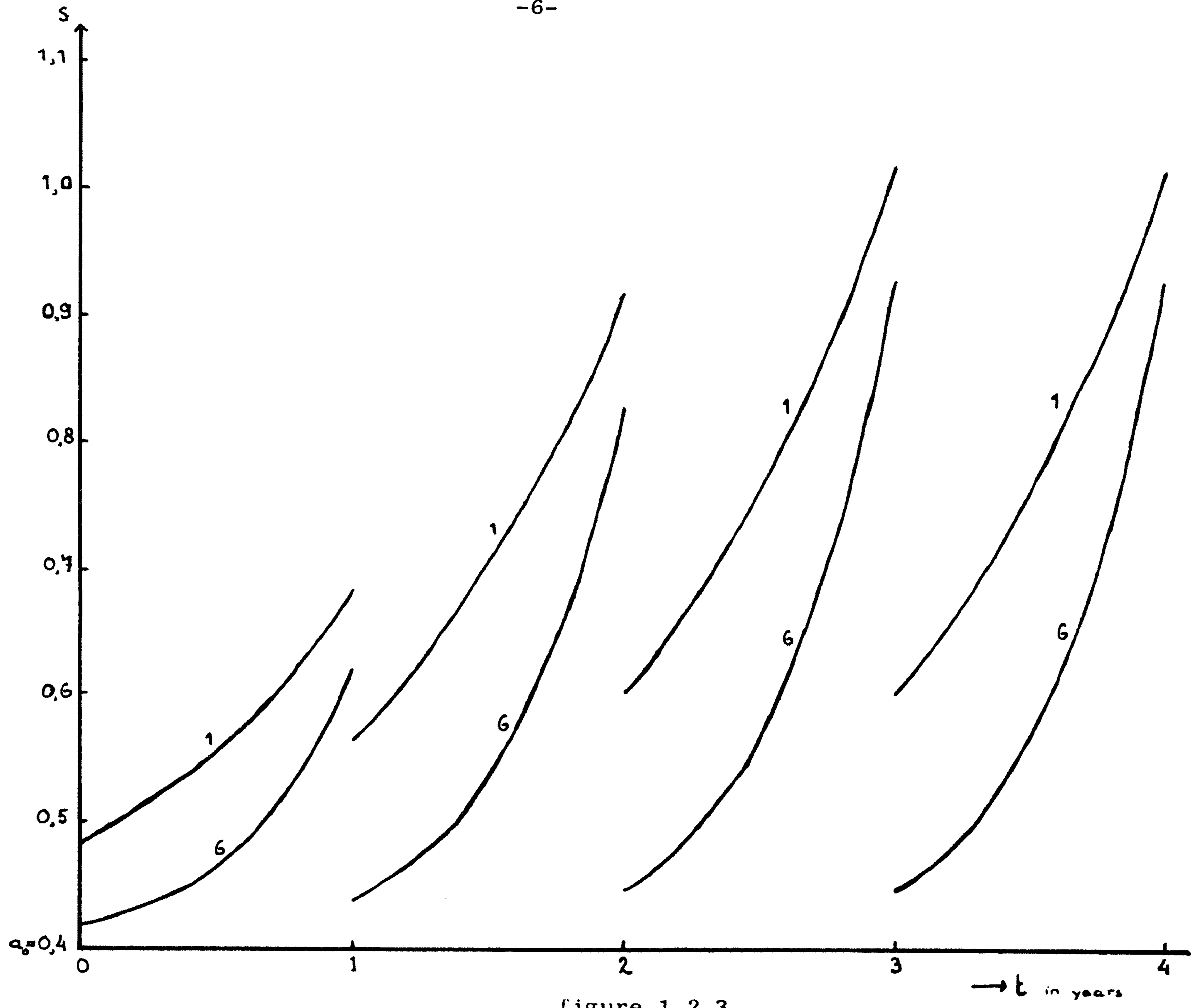


figure 1.2.3
The optimal strategy for two accident rates

2. Markov-programming

In problems of the type to which Markov-programming can be applied there is always a question of a physical system. In our case the system comprises the car and the accident insurance.

At each point of time t the system is in some state x . In the mathematical model the state x is represented by a point in a finite dimensional Cartesian space, called the state space X .

Apart from deterministic transformations the state of the system is subject to random transitions. Owing to these transitions the system performs a random walk through the state space X . If the decisionmaker does not intervene, such a walk is said to be a realisation of the natural process. A condition for application of Markov-programming is that for each initial state the underlying natural process can be described by a stationary strong Markov-process.

A stationary Markov-process is characterized by the following property: "Suppose the system will be in state x at a fixed time t then the probability of being in some set of states A at time $t+s$ depends only on A , x and s ". In formula:

$$(2.1) \quad P^{t+s}(A; x, t) = P^s(A; x, 0) = P^s(A; x).$$

In this paper a stationary Markov-process is called strong if the above relation remains true when t is not given beforehand but will be fixed by the random time the system enters an arbitrary but given closed set of states in X . In other words: states, assumed by the system before the state x is reached, are irrelevant for future transitions when x is completely specified. Further the distribution of the transition probabilities is independent of t , even if t is random in the sense we mentioned above.

In our motorist problem, the natural process results from the passage of time and the occurrence of accidents. The assumption that the number of accidents in a time interval T is Poisson distributed with constant parameter and the assumed independence between successive damages together imply that the natural process in our problem is a stationary strong Markov-process.

In general the decisionmaker will try to influence the natural process by interventions, basically a finite number in a finite interval. After such an intervention the system is transferred into some other state. Between interventions the system is subject to the natural process. For that reason the natural process has to be defined for each initial state. It is convenient to assume that at each point of time a decision is made. The decision will be primarily to decide whether to intervene or not and secondly which intervention to choose. In the case the decision is not to intervene we will speak of a null-decision. Once it is decided to intervene in some state x we will have to decide among the different possibilities which intervention is going to be effected. We shall assume that in every state x there exists a set $D(x)$ of possible decisions d . Mathematically a decision is defined by the probability distribution of the state into which the system is transferred (by the decision!). A null-decision in the state x is a probability distribution concentrated in x itself. In the motorist problem decisions lead to deterministic transitions. Consequently, these decisions are also defined by "concentrated" probability distributions but now in the new state. As soon as to every state a decision (including null-decisions) has been attached we have a strategy. Hence a strategy shall specify the set of states where the decisionmaker will intervene and in addition to this it determines the probability distribution of the state just after the intervention.

The resultant of the natural process and the transitions dictated by the strategy is called the decision-process. Under certain general conditions it can be proved that the decision-process is also a stationary strong Markov-process.

With regards to the decision-process in the state space a set of so-called transient states can be differentiated from one or more simple ergodic sets. The set of transient states has the property that with probability 1 the system will never return to this set once it has left it. Ergodic sets are characterized by the fact that once the system assumes a state of such a set it remains in that set forever with probability 1.

A simple ergodic set cannot be subdivided into disjunct ergodic sets.

In general a decomposition of the state space in a set of transient states and simple ergodic sets is not completely exhaustive.

The set of intervention states plays a prominent part in the decision-process. For a strategy z the set of intervention states will be denoted by A_z .

Let \underline{I}_n ($n = 1, 2, \dots$) be the sequence of future intervention states, if the strategy z is applied. The sequence \underline{I}_n ($n = 1, 2, \dots$) constitutes a stationary Markov-process with a discrete time parameter. The probability distribution of \underline{I}_n , given the initial state x , will be denoted by

$$(2.2) \quad p^{(n)}(A; z; x) \quad n = 1, 2, \dots \quad *)$$

Often it can be proved that the stationary distribution of the "intervention state \underline{I}_∞ " exists and is given by

$$(2.3) \quad \phi(A; z; x) = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n p^{(k)}(A; z; x).$$

It can also be proved that, if x_1 and x_2 are initial states from the same simple ergodic set, the stationary distributions are equal; i.e.

$$(2.4) \quad \phi(A; z; x_1) = \phi(A; z; x_2).$$

The optimal strategy has to be chosen according to some criterion. Which criterion has to be used? Let us consider a realization of the process. Such a walk w through the state space X may be represented by a point w in the space of all possible walks Ω . Applying strategy z during walk w we denote the costs in a time period T by $k_T(w; z)$. If $T \rightarrow \infty$, in most cases $\lim_{T \rightarrow \infty} k_T(w; z) = \infty$. Another disadvantage is that the walk w is not known in advance. By considering the average costs per time unit we can overcome these objections. For this criterion one can prove the following theorem:

*) A is some set of states in A_z .

This probability distribution can be extended to the whole space by taking

$$p^{(n)}(A; z; x) = p^{(n)}(A \cap A_z; z; x).$$

If w denotes a walk of the system, starting in x_0 , x_0 being a state of an simple ergodic set, then

$$(2.5) \quad \lim_{T \rightarrow \infty} \frac{k_T(w; z)}{T}$$

exists with probability 1 and is equal to

$$(2.6) \quad \frac{\int_{A_z} \phi(dI; z; x_0) k(I; z)}{\int_{A_z} \phi(dI; z; x_0) t(I; z)},$$

where $k(I; z)$ denotes the expected costs during- and $t(I; z)$ the expected length of- the time period between the intervention state I and the next intervention state assumed by the system. In this paper the criterion for optimality deals with the average costs per unit of time and is denoted by the function $r(z; x_0)$, where z is the applied strategy and x_0 is the initial state. If x_0 belongs to a simple ergodic set, the criterion function is defined by

$$(2.7) \quad r(z; x_0) = \frac{\int_{A_z} \phi(dI; z; x_0) k(I; z)}{\int_{A_z} \phi(dI; z; x_0) t(I; z)}.$$

For two states x_1 and x_2 of the same simple ergodic set we have [c.f. (2.4)]

$$(2.8) \quad r(z; x_1) = r(z; x_2).$$

If x_0 is a transient state the limit (2.5) still exists with probability 1 but is unknown until one of the simple ergodic sets has been entered; the limit (2.5) is equal to that of the entry state.

Consequently, if x_0 is a transient state, (2.5) is at x_0 a random variable.

Let us assume that there are m simple ergodic sets, denoted by E_r ($r = 1, 2, \dots, m$). It is easily verified that the probability of entering the ergodic set E_r , starting in x_0 , is given by $\phi(E_r; z; x_0)$.

If x_0 is a transient state the expected average costs per unit of time are given by

$$(2.9) \quad \sum_{r=1}^m \phi(E_r; z; x_0) r(z; e_r),$$

where e_r is some state in E_r [c.f. (2.8)].

Obviously, (2.9) is identical with

$$(2.10) \quad \int_{A_z} \phi(dy; z; x_0) \frac{\int_{A_z} \phi(dI; z; y) k(I; z)}{\int_{A_z} \phi(dI; z; y) t(I; z)}$$

We now define the criterion function $r(z; x_0)$ by

$$(2.11) \quad r(z; x_0) = \int_{A_z} \phi(dy; z; x_0) \frac{\int_{A_z} \phi(dI; z; y) k(I; z)}{\int_{A_z} \phi(dI; z; y) t(I; z)} .$$

This definition comprises (2.7).

A strategy is called optimal with respect to a class Z of strategies, if for each x

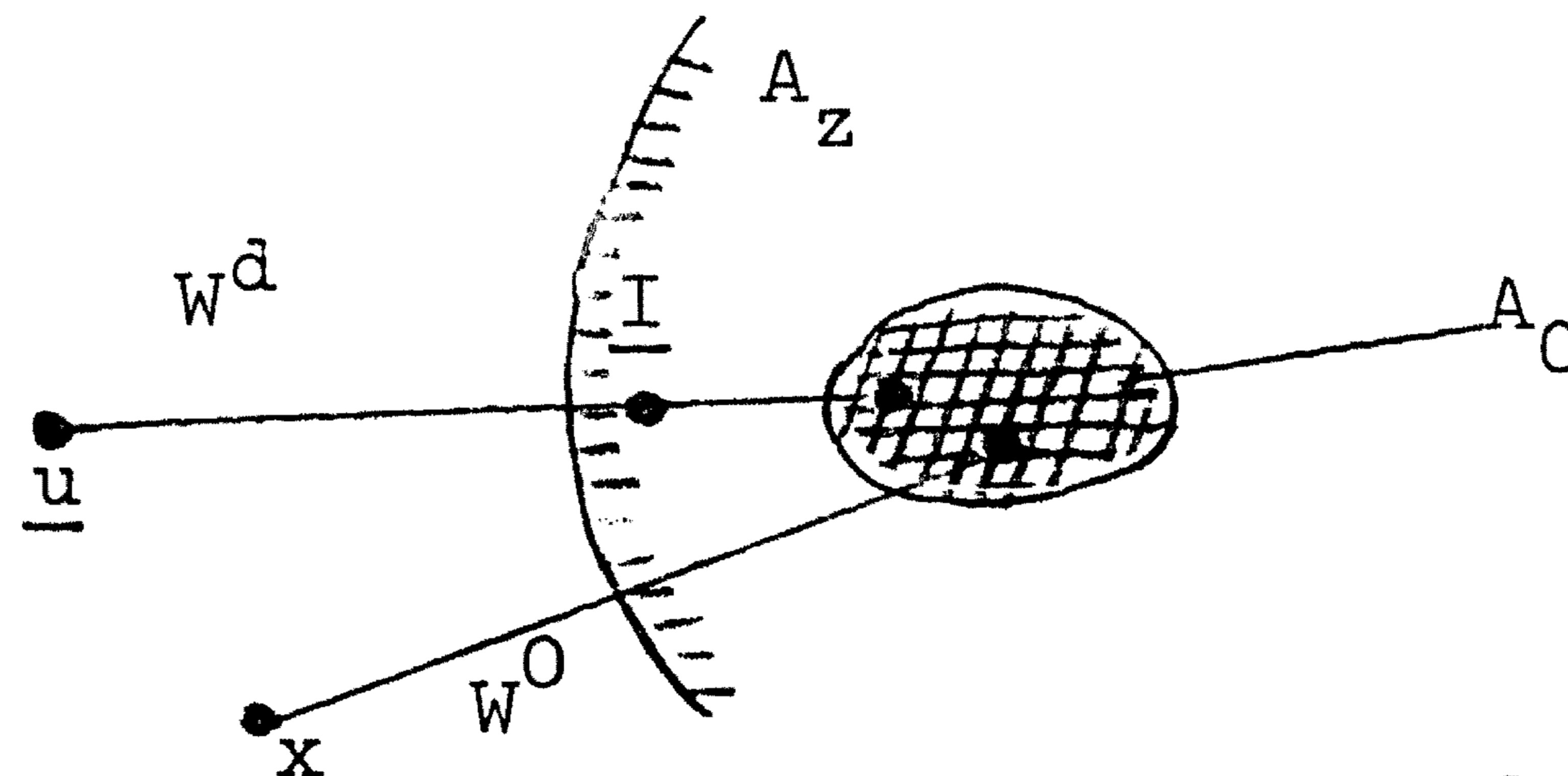
$$(2.12) \quad r(z_0; x) = \min_{z \in Z} r(z; x)$$

To determine the criterion function $r(z; x)$ we need to know the functions $k(I; z)$ and $t(I; z)$. These two functions still depend on the strategy z . We now show that it is possible to define the function $r(z; x)$ with the aid of two somewhat different functions $k(x; d)$ and $t(x; d)$. Both functions do not depend on the strategy applied, but apart from the state x only on the decision d made in x .

To this end we consider a set A_0 , not empty. The set A_0 consists of states where each strategy $z \in Z$ dictates an intervention. Hence, if A_z is the intervention set of an arbitrary strategy $z \in Z$,

$$(2.13) \quad A_z \supset A_0.$$

In this paper it is assumed that in the natural process from each initial state the set A_0 can be reached within a finite time with probability 1. For each state x and every decision $d \in D(x)$ we consider two walks denoted by W^0 and W^d . During W^0 the system will be subject to the natural process until a state of A_0 has been reached. During W^d the decision d transforms the system to the random state \underline{u} (with probability distribution d). From state \underline{u} the walk W^d will be subject to the natural process. It follows from (2.13) that for each strategy z the system will reach the set A_0 via or in an intervention state $\underline{I} \in A_z$ (figure 2.1).



Schematical representation of the walk W^d
and W^0 .

figure 2.1

Let for the walk W^0 $k_0(x)$ and $t_0(x)$ be the expected costs and the expected duration respectively. Let for the walk W^d $k_1(x; d)$ and $t_1(x; d)$ be the expected costs and the expected duration respectively. We now define the functions $k(x; d)$ and $t(x; d)$ to be the difference in expected costs and expected duration between the walks W^d and W^0 . In formula:

$$(2.14) \quad k(x; d) = k_1(x; d) - k_0(x).$$

$$(2.15) \quad t(x; d) = t_1(x; d) - t_0(x).$$

For null-decisions W^d and W^0 are identical and consequently:

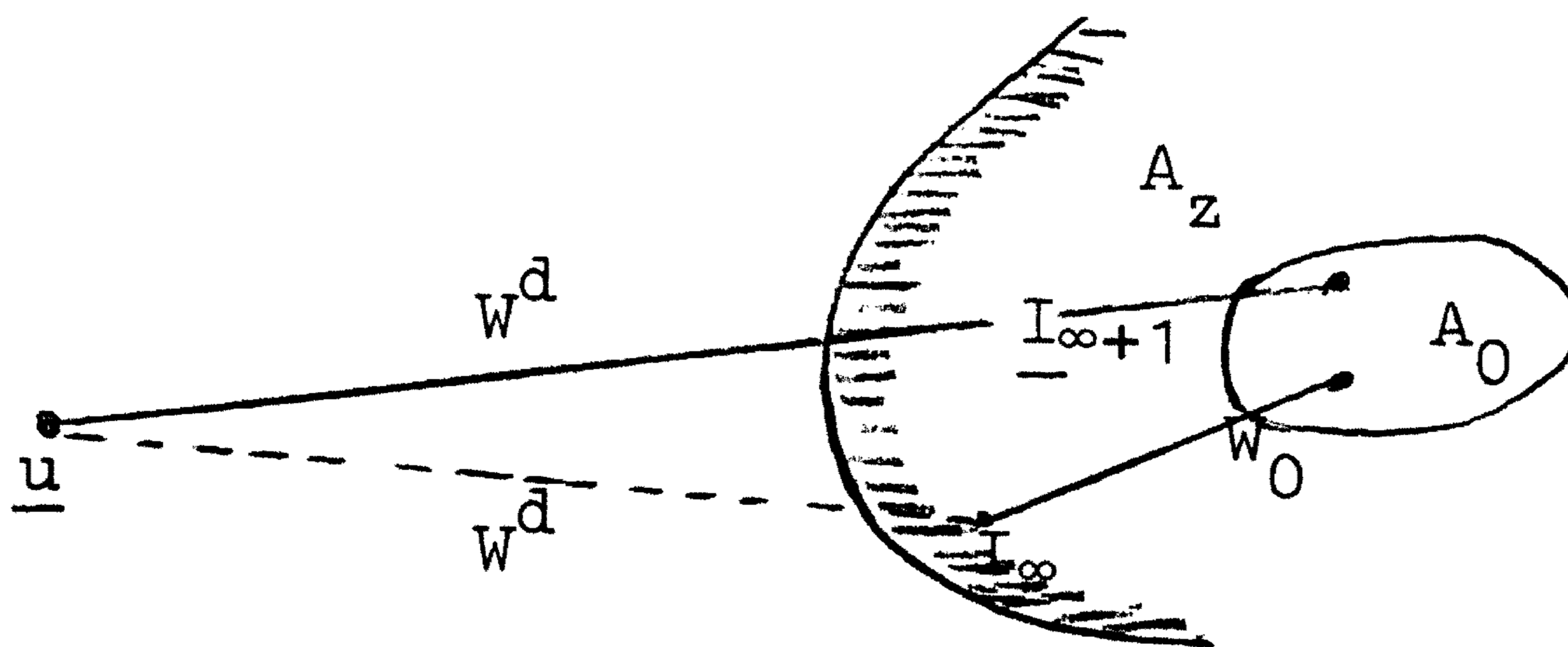
$$(2.16) \quad k(x; d) = 0.$$

$$(2.17) \quad t(x; d) = 0.$$

Note that the functions $k(x; d)$ and $t(x; d)$ do not depend on a particular strategy.

It follows from their definitions that $k(I_\infty; z)$ and $t(I_\infty; z)$ are identical to the expected cost and duration of the part of the walk W^d between the intervention states \underline{I}_∞ and $\underline{I}_{\infty+1}$ (see figure 2.2) ^{*)}

Both states \underline{I}_∞ and $\underline{I}_{\infty+1}$ are distributed according to the same limiting distribution $\phi(A; z; y)$. Hence, with respect to the initial state y , the expected costs (duration) from \underline{I}_∞ and those (that) from $\underline{I}_{\infty+1}$ to A_0 are (is) equal.



Schematical representation of the walk W^d and W^0 .

figure 2.2

This implies:

$$(2.18) \quad \int_{A_z} \phi(dI_\infty; z; y) k(I_\infty; z) = \int_{A_z} \phi(dI_\infty; z; y) \cdot k(I_\infty; z(I_\infty))$$

$$(2.19) \quad \int_{A_z} \phi(dI_\infty; z; y) t(I_\infty; z) = \int_{A_z} \phi(dI_\infty; z; y) \cdot t(I_\infty; z(I_\infty))$$

where $z(I_\infty)$ denotes the decision dictated by the strategy in state I_∞ . It follows from (2.18) and (2.19) that instead of (2.11) the criterion function $r(z; x_0)$ can also be defined by

^{*)} What is in a notation?

$$(2.20) \quad r(z; x_0) = \int_{A_z} \phi(dy; z; x_0) \frac{\int_{A_z} \phi(dI; z; y) k(I; z(I))}{\int_{A_z} \phi(dI; z; y) t(I; z(I))}$$

It follows from (2.8) that the criterion function $r(z; x)$ does not pronounce upon the most profitable initial state. We like a preference function $\hat{c}(z; x)$ having the property that for two states x_1 and x_2 in the same simple ergodic set the difference in total expected costs is given by

$$(2.21) \quad \hat{c}(z; x_2) - \hat{c}(z; x_1).$$

If the state space can be decomposed in a set of transient states and m disjunct simple ergodic sets E_j and if states e_j are arbitrarily chosen states in E_j , let a function $c(z; x)$ be defined by

$$(2.22) \quad c(z; x) = k(x; z(x)) - r(z; x) t(z; z(x)) + \int_{A_z} p^{(1)}(dI; z; x) c(z; I).$$

$$(2.23) \quad c(z; e_j) = 0 \quad j = 1, 2, \dots, m.$$

It can be proved that a preference function $\hat{c}(z; x)$ of the type mentioned above is defined by

$$(2.24) \quad \hat{c}(z; x) = c(z; x) + k_0(x) - r(z; x) t_0(x).$$

Presently it will appear that the functions $r(z; x)$ and $c(z; x)$ are all-important. They can be obtained simultaneously by solving the following functional equations:

$$(2.25) \quad r(z; x) = \int_{A_z} p^{(1)}(dI; z; x) r(z; I),$$

$$(2.22) \quad c(z; x) = k(x; z(x)) - r(z; x) t(x; z(x)) + \int_{A_z} p^{(1)}(dI; z; x) c(z; I),$$

$$(2.23) \quad c(z; e_j) = 0 \quad j = 1, 2, \dots, m,$$

where e_j is an arbitrarily chosen state in the simple ergodic set E_j .
 Summarizing: Independent of the strategy to apply the functions $k(x; d)$ and $t(x; d)$ can be defined. As soon as a strategy z and a set of points $e_j \in E_j$ ($j = 1, \dots, m$) have been chosen, the functions $r(z; x)$ and $c(z; x)$ are unambiguously defined by (2.25), (2.22) and (2.23). Note that, if $x \in A_z$ [c.f. (2.16) and (2.17)] we have

$$(2.26) \quad c(z; x) = \int_{A_z} p^{(1)}(dI; z; x) c(z; I).$$

Now the properties of the optimal strategy will be outlined. Further based on these properties an iteration procedure will be constructed.

The iteration procedure yields a sequence of strategies $z^{(i)}$ ($i = 1, 2, \dots$) of which, under certain conditions, the following interesting properties can be proved:

$$(2.27) \quad \begin{aligned} \text{a)} \quad & r(z^{(i)}; x) \geq r(z^{(i+1)}; x) \\ \text{b)} \quad & \lim_{i \rightarrow \infty} r(z^{(i)}; x) = \min_{z \in Z} r(z; x) \end{aligned}$$

where Z is the class of all admissible strategies. Proofs and conditions are given in [1] and will be omitted here. We will restrict ourselves to some definitions and to a glance at the procedure. We start with a bare catalogue of definitions. Suppose that in the initial state x the decision d is made. If decision d transforms the system into the random state \underline{u} and if after d the strategy z is applied, then functions $r(d.z; x)$ and $c(d.z; x)$ are defined as follows:

$$(2.28) \quad r(d.z; x) \stackrel{\text{def}}{=} E\{r(z; \underline{u}) | d\}$$

$$(2.29) \quad c(d.z; x) \stackrel{\text{def}}{=} k(x; d) - r(d.z; x) t(x; d) + E\{c(z; \underline{u}) | d\},$$

We easily verify that for both null-decision and $d = z(x)$, we have

$$(2.30) \quad r(d.z; x) = r(z; x)$$

$$(2.31) \quad c(d.z; x) = c(z; x)$$

Suppose the system is now in x and let \underline{v} be the first state in a closed set A assumed by the system, then the functions $r(A.z; x)$ and $c(A.z; x)$ are defined by:

$$(2.32) \quad r(A.z; x) \stackrel{\text{def}}{=} E\{r(z; \underline{v}) | x; A\}$$

$$(2.33) \quad c(A.z; x) \stackrel{\text{def}}{=} E\{c(z; \underline{v}) | x; A\}.$$

Besides we define the class K_z of all closed sets $A \supset A_0$ satisfying:

$$(2.34) \quad \{x | r(A.z; x) < r(z; x)\} \cup \\ \cup \{x | r(A.z; x) = r(z; x); c(A.z; x) \leq c(z; x)\} = X,$$

where X denotes the state space. We easily verify that $A_z \in K_z$.

Finally we define the following subsets:

$$(2.35) \quad D_z(x) \stackrel{\text{def}}{=} \{d | d \in D(x); r(d.z; x) = \min_{d^* \in D(x)} r(d^*.z; x)\}$$

and

$$(2.36) \quad A'_z = \bigcap_{A \in K_z} A.$$

In order to gain an insight into the principle of solution, we consider the following problem:

Suppose a decisionmaker has to make his decisions in accordance with a strategy z . In the initial state however he is free to choose a decision d . Which decision is the most profitable?

The decisionmaker certainly looks for that particular decision which minimizes the expected average costs per unit of time. Each fall in these costs leads to an infinite saving in an infinite period of time.

If he selects in x a decision d , the expected average costs per unit of time are given by $r(d.z; x)$. So the decisionmaker has to solve the problem [c.f. (2.28)]

$$(2.37) \quad \min_{d \in D(x)} r(d.z; x).$$

With respect to the effect of the initial state in the total expected costs we introduced a preference function $\hat{c}(z; x)$; this function can be defined by (2.24). With respect to the effect of a decision d in the initial state x we need a preference function $\hat{c}(d.z; x)$ such that the difference in the total expected costs can be expressed by

$$(2.38) \quad \hat{c}(d.z; x) - \hat{c}(z; x).$$

This difference has to be attributed to the decision d . It can be proved that the difference is also measured by [c.f. (2.29)]:

$$(2.39) \quad c(d.z; x) - c(z; x)$$

A possible fall in costs will in general be finite.

Consequently, if more than one decision d minimizes $r(d.z; x)$, the decisionmaker can use his freedom by minimizing (2.39) with respect to these equivalent decisions; or, what is the same, by minimizing the function $c(d.z; x)$ with respect to $d \in D_z(x)$. [c.f. (2.35)].

Summarizing: Essentially the decisionmaker has to solve the following two problems.

- 1) To minimize the d -function $r(d.z; x)$ with respect to $d \in D(x)$.
- 2) To minimize the d -function $c(d.z; x)$ with respect to $d \in D_z(x)$.

If these two problems have been solved for each x , then to each x a minimizing decision d can be added. If $z(x)$ belongs to $D_z(x)$ and also minimizes $c(d.z; x)$, let the decision $z(x)$ be chosen. The relation between states and decisions is nothing else than a strategy. Let this strategy be denoted by z_1 . The following important result can now be proved:

$$(2.40) \quad r(z_1; x) \leq r(z; x)$$

So the solution of the decisionproblem mentioned above also offers a new strategy; a strategy being at least as good as strategy z . Let us examine the new strategy. It follows from (2.30) and (2.31) that all intervention states of strategy z are also intervention states of strategy z_1 .

Hence

$$(2.41) \quad A_{z_1} \supset A_z.$$

In other words in the initial state the decisionmaker can change but not defer the intervention dictated by the original strategy z .

This important result leads us to a second decisionproblem.. Suppose that the decisionmaker has to make his decisions in accordance with a strategy z . But he is allowed to determine the point of time whereupon the strategy comes into operation. This will be done by choosing a closed set A ; the strategy comes into operation at the moment the system enters the set A for the first time. Which set is the most profitable? The decisionmaker certainly looks for a delay that minimizes the expected average costs per unit of time. This implies that sets A will be considered which satisfy for each x [c.f. (2.32)]

$$(2.42) \quad r(A.z; x) < r(z; x).$$

Again each fall in the average costs leads to an infinite saving in an infinite period of time. With respect to the effect of a delay in the total expected costs we need a preference function $\hat{c}(A.z; x)$ such that the difference in expected costs, measured by

$$(2.43) \quad \hat{c}(A.z; x) - \hat{c}(z; x),$$

can be attributed to the delay. It can be proved that this difference is also given by [c.f. (2.33)]

$$(2.44) \quad c(A.z; x) - c(z; x).$$

Consequently, sets A which satisfy for each x

$$(2.45) \quad r(A.z; x) = r(z; x)$$

and

$$(2.46) \quad c(A.z; x) \leq c(z; x)$$

will also be considered.

Summarizing: Closed sets A having the property [c.f. (2.34)]

$$(2.47) \quad \begin{aligned} X &= \{x | r(A.z; x) < r(z; x)\} \cup \{x | r(A.z; x) = \\ &= r(z; x); c(A.z; x) \leq c(z; x)\} \end{aligned}$$

will be considered. If K_z is the class of all closed sets A satisfying (2.47) and if

$$(2.48) \quad A'_z = \bigcap_{A \in K_z} A$$

belongs to K_z , then obviously the set A'_z is the solution of the second decision problem.

It can be proved that the strategy z_2 defined by

$$(2.49) \quad z_2(x) = \begin{cases} z(x) & \text{if } x \in A'_z \cap A_z \\ \text{null-decisions} & \text{otherwise} \end{cases}$$

satisfies:

$$(2.50) \quad r(z_2; x) \leq r(z; x).$$

From the solutions of the two decision problems considered here above we now deduce the following properties of the optimal strategy z_0 :

$$(2.51) \quad \min_{d \in D(x)} r(d.z_0; x) = r(z_0; x)$$

$$(2.52) \quad \min_{d \in D_{z_0}(x)} c(d.z_0; x) = c(z_0; x)$$

$$(2.53) \quad A'_{z_0} = A_{z_0}.$$

The iteration procedure runs as follows:

Preparatory part

Determine the $(x; d)$ -functions $k(x; d)$ and $t(x; d)$.

Iterative approach

Let $z^{(n-1)}$ be the strategy obtained at the $(n-1)$ th cycle of the iteration procedure.

- 1) Determine the function $r(z^{(n-1)}; x)$ and $c(z^{(n-1)}; x)$ by solving the functional equations (2.25), (2.22) and (2.23).
- 2) a) Determine the functions $r(d, z^{(n-1)}; x)$ and $c(d, z^{(n-1)}; x)$ by using the relations (2.28) and (2.29).
 b) Determine for each $x \in X$ the subset of minimizing decisions $D_z^{(n-1)}(x)$.
 c) Minimize for each $x \in X$ the d-function $c(d, x; z)$, subject to $d \in D_z^{(n-1)}(x)$.
 d) Add to each state x a solution of c). If $z^{n-1}(x)$ is a solution of c), this decision will be added to the state x . [This instruction has been made in order to advance the convergence of the sequence of strategies $\{z^{(1)}, z^{(2)}, \dots\}$].

As soon as the operation d) has been performed a new strategy $z_1^{(n-1)}$ has been constructed.

- 3) Determine the functions $r(z_1^{(n-1)}; x)$ and $c(z_1^{(n-1)}; x)$ by using the functional equations (2.25), (2.22) and (2.23).

- 4) Determine the subset $A'_{z_1^{(n-1)}}$. The new strategy $z^{(n)}$ is given by

$$(2.27) \quad z^{(n)}(x) = \begin{cases} z_1^{(n-1)}(x) & \text{if } x \in A'_{z_1^{(n-1)}} \\ \text{null-decision} & \text{otherwise} \end{cases}$$

End of the n th cycle.

The functions $r(z^{(n-1)}; c)$ and $c(z^{(n-1)}; x)$ are determined by functional equations. If these equations cannot be solved analytically they often can be solved numerically by Monte Carlo methods.

The way in which the set $A'_{z_1}(n-1)$ can be determined depends heavily on the structure of the decision problem considered. In the boundary points of the minimizing set A'_z it will often be indifferent whether to intervene or not. In the motorist problem, by example, we will see that this property leads to a differential equation for the optimal boundary of the set of states in which claims should be suppressed.

3. Application to the motorists problem

In this section it is shown how the motorist problem can be solved with Markov-programming.

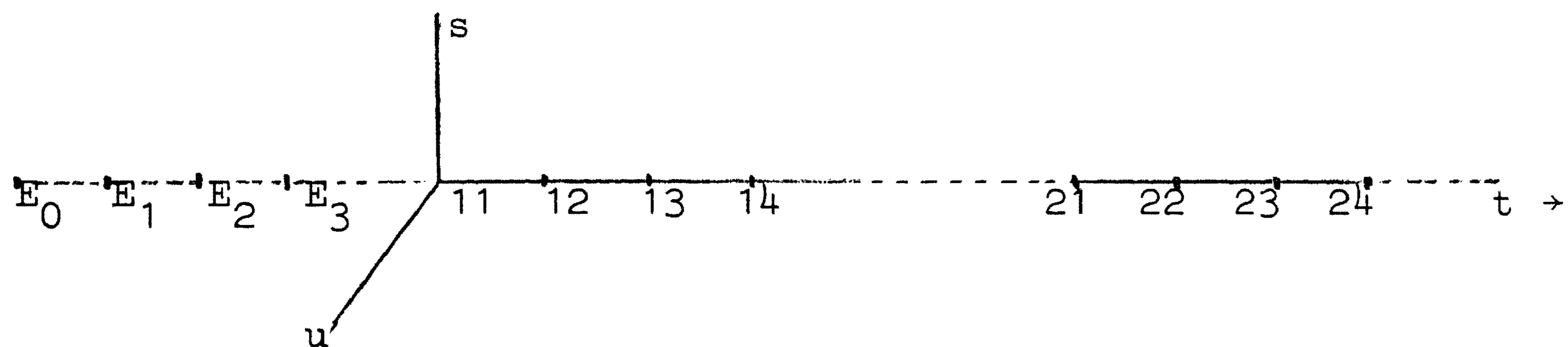
Primarily we shall have to define in detail the state space, the natural process, the set of intervention states and the set A_0 . Secondly we shall determine the k - and t - functions and finally the c - functions from which the optimal strategy is established.

3.1 Definition of the state space

At each point of time the following information will be of interest:

- (1) whether an eventual damage is covered or not.
- (2) whether an accident happens or not.
- (3) the amount of the last paid premium E_i , $i = 0, 1, 2, 3$.
- (4) the date and time of the day considered.
- (5) the extent of the damage.
- (6) whether a damage has been claimed since the last payment of premium or not.

The following state space is suggested:



The state space

fig. 3.1

At the t -axis we distinguish:

- a) Four points, namely E_0 , E_1 , E_2 and E_3 . In these states the corresponding premium has to be paid (January 1); damages are no longer covered by insurance.
- b) Four intervals of one year ^{*}), namely $1i \leq t < 1i + 1$ ($i = 1, 2, 3, 4$).

*⁾ $1i = 12$, if $i = 2$
 $1i = 13$, if $i = 3$.

The t -component of the state runs through $1i \leq t < 1i + 1$, if and only if the last premium paid was E_{i-1} , one or more damages have been claimed that year and the coming losses are still covered by insurance.

- c) Four intervals of one year, namely $2i \leq t < 2i + 1$ ($i = 1, 2, 3, 4$).
The t -component of the state runs through $2i \leq t < 2i + 1$, if and only if the last premium paid was E_{i-1} , no damages have been claimed that year and the coming damages are still covered by insurance.

The s-variabele is zero unless at least one damage has been claimed that year and moreover the coming damages are still covered by insurance. In that case the s -component denotes the extent of the last claim.

The u-variabele is zero unless at least one damage has been claimed that year and moreover the coming damages are still covered by insurance. In that case the u -component denotes the time elapsed since the first claim that year.

Note that the s - and u -components of the state can only be different from zero if $1i \leq t < 1i + 1$ ($i = 1, 2, 3, 4$). Consequently the state space consists of

- a) 4 points E_0, E_1, E_2, E_3 .
- b) a 3-dimensional subspace (t, s, u) with $11 \leq t < 15$.
- c) 1-dimensional intervals $2i \leq t < 2i + 1$ ($i = 1, 2, 3, 4$).

We are now in the position to describe the natural process. This process can start in each state of the state space. In accordance with the premium paid the system runs through one of the time-intervals $2i \leq t < 2i + 1$ ($i = 1, 2, 3, 4$), if no damage has been claimed that year. If no accident will happen during the rest of the year, at the end of the year the system is transferred to E_i . Since in the natural process no premiums are paid the system will stay there for ever. ^{*}).

*.) In the natural process no premiums will be paid. On the other hand the system can start in such a state that losses are covered by insurance for some time. This is no contradiction! It would be a contradiction if we had said: In the natural process no premiums have been paid.

However, if at t' ($t' < 2i + 1$) an accident occurs the system is transferred to $(t' - 10, s', 0)$, where s' denotes the loss incurred. Since during the natural process, irrespective of their extents, all losses are claimed, the system will stay in the 3-dimensional part of the state space for the rest of the year. From now on the u -component is increasing with time. The s -component will only change if a second, third etc. accident happens. At the end of the year the system is transferred to E_0 where it stays for ever.

If in the initial state (t, s, u) a damage has already been claimed that year and coming damages are still covered by insurance the description of the development in the state of the system is similar to that of the final part of a walk considered here above.

We now consider the decision-mechanism.

- a) In the points E_0, E_1, E_2 and E_3 two decisions are feasible: the null-decision and the decision that involves the payment of the premium E_i . Respective transformations are $E_i \rightarrow E_i$ and $E_i \rightarrow (2i+1, 0, 0)$.
- b) In the 3-dimensional space (t, s, u) only null-decisions are feasible, unless $u = 0$. If $u = 0$ an accident just occurs and consequently the decisionmaker can suppress the claim if he wants. In that case the system is transferred back to $(t + 10, 0, 0)$. Note that a claim corresponds with a null decision. This is in accordance with the fact that in the natural process all damages are claimed.

Since the u -component denotes the time elapsed since the first claim, it may happen that an accident occurs when $u > 0$. The decision not to claim the damage is of course a bad decision and for that reason it is considered to be infeasible.

- c) To states satisfying $2i \leq t < 2i + 1$ ($i = 1, 2, 3, 4$), only null-decisions are added.

In figure 3.2 states have been marked with more than one feasible decision.

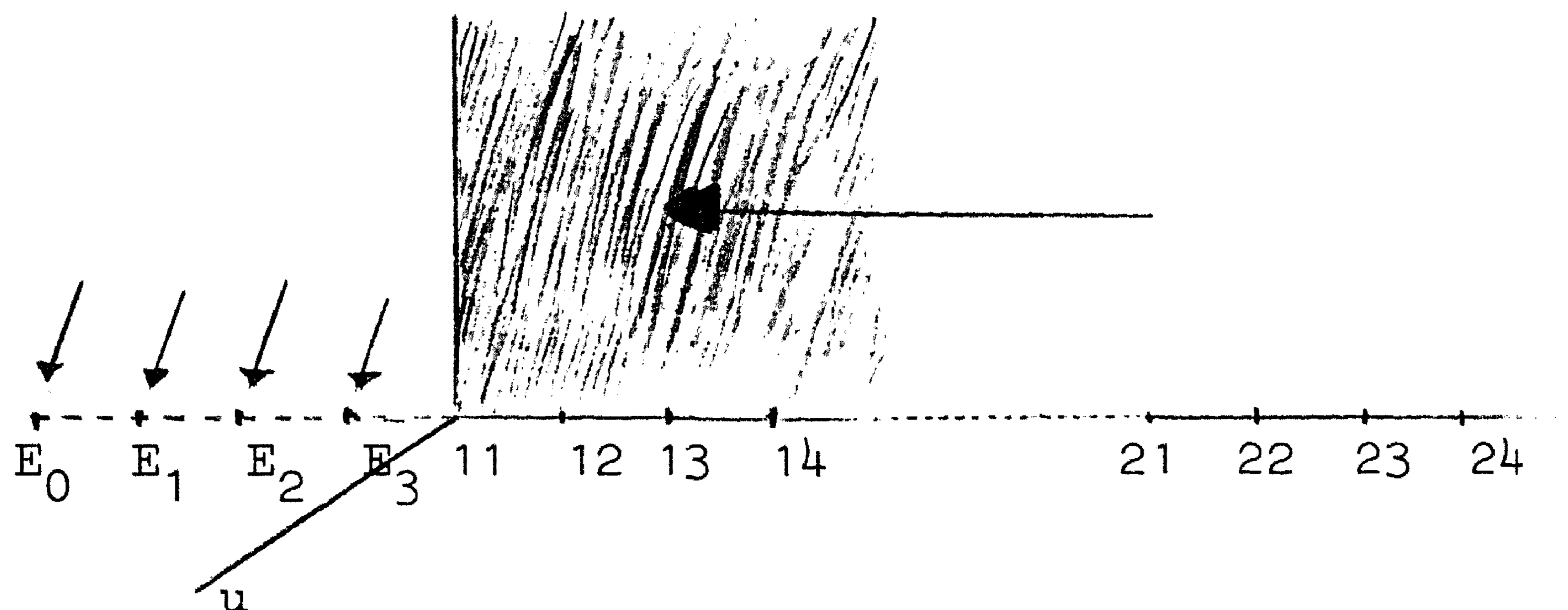


fig. 3.2

States with more than one feasible decision

From now on we shall only consider strategies which dictate payment of premium in the states E_i , $i = 0, 1, 2, 3$. Consequently these states may be chosen as elements of the set A_0 . We may add if we want the set of states (t, s, u) for which $1i \leq t < 1i + 1$, $s \leq a_0$ and $u = 0$, because every possible strategy will dictate a suppression of the claim in this set. However restriction of the set A_0 to the states E_i , $i = 0, 1, 2, 3$, leads to more simple expressions for the $k(x; d)$ and $t(x; d)$ functions. So we choose:

$$(3.1.1) \quad A_0 = \bigcup_{i=0}^3 E_i.$$

3.2 The determination of the $k(x; d)$ and $t(x; d)$ functions

Let us consider the intervention which results from an accident occurring τ time units after paying premium E_{i-1} and suppose that no damage has been claimed during the time interval $[0, \tau)$.

After the accident has occurred the system will be in state $x = (1i + \tau, s, 0)$ and the decision not to claim transfers it back into state $2i + \tau$, hence $d = 2i + \tau$. We note first that the W^d walk after the decision $d = 2i + \tau$ and the W^0 walk are both subject to the natural process until at the end of the year the set A_0 is reached. The walk W^d terminates in state E_i

and the W^0 walk in E_0 . In both cases however the expected costs and the expected duration are equal. The only difference in expected costs is given by the decision costs of no claiming which amount $s - a_0$ if $s > a_0$ and 0 if $s \leq a_0$. Consequently we have

$$(3.2.1) \quad t(x;d) = t_1(x;d) - t_0(x) = 0$$

$$(3.2.2) \quad k(x;d) = k_1(x;d) - k_0(x) = \begin{cases} s - a_0 & s > a_0 \\ 0 & s \leq a_0 \end{cases}$$

Secondly we need to consider the intervention of the payment of premium. The system will be in state E_{i-1} and the decision of paying premium transfers it to state $2i$. Because $E_{i-1} \in A_0$ the expected duration as well as the expected costs in the walk are zero.

During the W^d walk after the payment of premium E_{i-1} the system is subject to the natural process until at the end of the year the set A_0 is reached by means of state E_i . Hence the expected duration will be one year. Because in the natural process all damages are claimed we have for the expected costs per accident, denoted by $k(a_0)$,

$$(3.2.3) \quad k(a_0) = \int_0^{a_0} s \, dF(s) + a_0 \int_{a_0}^{\infty} dF(s).$$

Consequently the expected costs in the natural process during the interval $[0,1]$ are $\lambda k(a_0)$.

$$(3.2.4) \quad t(x;d) = 1$$

$$(3.2.5) \quad k(x;d) = E_{i-1} + \lambda k(a_0).$$

3.3 Determination of the optimal strategy

It is easily verified for all strategies $z \in Z$ the Markov process in A_z has only one simple ergodic set. Consequently for every strategy z and feasible decision d , we have

$$(3.3.1) \quad r(d.z; x) = r(z;x) = r(z).$$

Hence it is sufficient to consider the function $c(z;x)$, defined by the functional equation

$$(3.3.2) \quad c(z;x) = k(x;z(x)) - r(z) t(x;z(x)) + \int_{A_z} p^{(1)}(dI;z;x) c(z;I).$$

In order to obtain a unique solution we put

$$(3.3.3) \quad c(z;E_0) = 0.$$

Let us first consider the solution in state $x = (t,s,0)$ with $t = 1i + r$ and $s > a_0$. Suppose we are applying an arbitrary strategy z . If z dictates to claim (null decision!) in x then the next intervention state is E_0 . So we have

$$(3.3.4) \quad c(z;x) = c(z;E_0) = 0.$$

If we decide not to claim and future decisions are taken in accordance with strategy z , then the function $c(d.z;x)$ is given by

$$(3.3.5) \quad c(d.z; x) = s - a_0 + c(z;t + 10).$$

From now on we consider only the optimal strategy z_0 . Let the boundary of A_{z_0} be given by the function $s = s(t)$. For z_0 holds:

$$(2.52) \quad c(z_0;x) = \min_{d \in D(x)} c(d.z_0; x).$$

For $a_0 < s \leq s(t)$ it will be profitable not to claim; thus $d = z_0(x) = t + 10$. It follows from (3.3.4) and (2.52)

$$(3.3.6) \quad c(z_0;x) \leq c(z_0;E_0) = 0.$$

According to the fact that $c(d.z_0;x)$ is a linear function of s it will be indifferent on the boundary $s(t)$ of A_{z_0} whether to claim or not, hence for $s = s(t)$ we have:

$$(3.3.7) \quad c(z_0;(t,s,0)) = 0.$$

From (3.3.5) and (3.3.7) it follows that

$$(3.3.8) \quad c(z_0;(t,s(t), 0)) = s(t) - a_0 + c(z_0;t + 10) = 0.$$

Consequently:

$$(3.3.9) \quad c(z_0; t + 10) = a_0 - s(t)$$

and in accordance with (3.3.5), by virtue of (3.3.9), for $s > a_0$:

$$(3.3.10) \quad c(z_0; (t, s, 0)) = s - s(t).$$

For $s \leq a_0$ by (3.2.1) and (3.2.2):

$$(3.3.11) \quad c(z_0; (t, s, 0)) = c(z_0; t + 10) = a_0 - s(t).$$

Furthermore holds for the states $E_i, i = 1, 2, 3$:

$$(3.3.12) \quad c(z_0; E_i) = \lim_{t \uparrow 2i+1} c(z_0; t)$$

or

$$(3.3.13) \quad c(z_0; E_i) = a_0 - \lim_{t \uparrow 2i+1} s(t).$$

Summarizing:

$$c(z_0; x) =$$

$$\begin{array}{l} \left| \begin{array}{ll} 0, & x \in E_0 \cup \{11 \leq t < 15, s > s(t), u = 0\} \cup \\ & \{11 \leq t < 15, s \geq 0, u > 0\} \\ a_0 - \lim_{t \uparrow 2i+1} s(t), & x \in \bigcup_{i=1}^3 E_i \\ a_0 - s(t), & x \in \{11 \leq t < 15, s \leq a_0, u = 0\} \\ s - s(t), & x \in \{11 \leq t < 15, a_0 < s \leq s(t), u = 0\} \\ a_0 - s(t), & x \in \{21 \leq t < 25\} \end{array} \right. \end{array}$$

From the functional equation (3.3.2) it follows for $E_{i-1}, i = 1, 2, 3, 4$:

$$(3.3.15) \quad c(z_0; E_{i-1}) = k(E_{i-1}; 2i) - r(z_0) t(E_{i-1}; 2i) + c(z_0; 2i)$$

or

$$(3.3.16) \quad c(z_0; E_{i-1}) - c(z_0; 2i) = E_{i-1} + \lambda k(a_0) - r(z_0).$$

Using the relations (3.3.14) for $x = E_{i-1}$, $i = 1, 2, 3, 4$

$$(3.3.17) \quad s(11) = E_0 + \lambda k(a_0) - r(z_0) + a_0$$

$$(3.3.18) \quad s(1i) = \lim_{t \uparrow 1i} s(t) + E_{i-1} + \lambda k(a_0) - r(z_0), \quad i = 2, 3, 4$$

$$(3.3.19) \quad \lim_{t \uparrow 14} s(t) = \lim_{t \uparrow 15} s(t)$$

For $x = (t, s, 0)$ with $s > a_0$ and $t = 1i + \tau$ follows from (3.3.2)

$$(3.3.20) \quad c(z_0; (t, s, 0)) = k((t, s, 0); t + 10) - r(z_0) t((t, s, 0); t + 10) \\ + \int_{1i+1-t}^{\infty} c(z_0; E_i) \lambda e^{-\lambda \tau_1} d\tau_1 \\ + \int_0^{1i+1-t} \lambda e^{-\lambda \tau_1} d\tau_1 \int_0^{s(t+\tau_1)} c(z_0; (t+\tau_1, y, 0)) dF(y) \\ + \int_0^{1i+1-t} \lambda e^{-\lambda \tau_1} d\tau_1 \int_{s(t+\tau_1)}^{\infty} c(z_0; E_0) dF(y)$$

According to (3.2.1), (3.2.2) and (3.3.14):

$$(3.3.21) \quad c(z_0; (t, s, 0)) = s - a_0 + e^{-\lambda(1i+1-t)} \left(a_0 - \lim_{t \uparrow 1i+1} s(t) \right) \\ + \int_0^{1i+1-t} \lambda e^{-\lambda \tau_1} d\tau_1 \int_{a_0}^{s(t+\tau_1)} (y - s(t+\tau_1)) dF(y) \\ + \int_0^{1i+1-t} \lambda e^{-\lambda \tau_1} d\tau_1 \int_0^{a_0} (a_0 - s(t+\tau_1)) dF(y)$$

If $\tau_2 = t + \tau_1$, then

$$(3.3.22) \quad c(z_0; (t, s, 0)) = s - a_0 + e^{-\lambda(1i+1-t)} \left(a_0 - \lim_{t \uparrow 1i+1} s(t) \right)$$

$$\begin{aligned}
& + e^{\lambda t} \int_t^{l_{i+1}} \lambda e^{-\lambda \tau_2} d\tau_2 \int_{a_0}^{s(\tau_2)} (y - s(\tau_2)) dF(y) \\
& + e^{\lambda t} \int_t^{l_{i+1}} \lambda e^{-\lambda \tau_2} F(a_0) (a_0 - s(\tau_2)) d\tau_2.
\end{aligned}$$

After substituting $s = s(t)$ and using (3.3.7) the differentiation of (3.3.22) with respect to t leads to

$$(3.3.23) \quad \frac{d s(t)}{dt} = \lambda \int_{a_0}^{\infty} (y - a_0) dF(y) - \lambda \int_{s(t)}^{\infty} (y - s(t)) dF(y).$$

By partial integration this equation can also be written in the more simple form

$$(3.3.24) \quad \frac{d s(t)}{dt} = \lambda \int_{a_0}^{s(t)} (1 - F(y)) dy.$$

Except for a translation in the t direction the boundary $s(t)$ is determined by (3.3.24). In other words the boundary of A_{z_0} for $i = 1, 2, 3, 4$ are in the t -direction translated parts of one of the curves satisfying (3.3.24). The location of each part on this curve has to be determined from the relations (3.3.17) through (3.3.19). We will now show that this is possible.

Suppose that $r(z_0)$ is known, than $s(11)$ is solved from (3.3.17). From the curve $s = s(t)$ we deduce $\lim_{t \uparrow 12} s(t)$. From (3.3.18) for $i = 2$ we obtain $s(12)$. Similarly we can compute $\lim_{t \uparrow 13} s(t)$, $s(13)$, $\lim_{t \uparrow 14} s(t)$, $s(14)$ and $\lim_{t \uparrow 15} s(t)$.

This implies that if $r(z_0)$ is known, the optimal strategy is completely specified. As $r(z_0)$ is not known its value is determined by equation (3.3.19).

It should be noted that the differential equation (3.3.24) has an analytical solution in the case the damage per accident is exponentially distributed. We have then for $F(s) = 1 - e^{-\mu s}$

$$(3.3.25) \quad \frac{d s(t)}{dt} = \frac{\lambda}{\mu} e^{-\mu a_0} (1 - e^{-\mu(s(t)-a_0)})$$

which leads to

$$(3.3.26) \quad s(t) = a_0 + \frac{1}{\mu} \ln \{1 + e^{\lambda(t+c_i)} e^{-\mu a_0}\} \quad \begin{array}{l} 1i \leq t < 1i + 1 \\ i = 1,2,3,4 \end{array}$$

where the $c_i, i = 1,2,3,4$ are arbitrary integration constants. Note that (3.3.26) determines a curve except for a translation in the t -direction. If the distribution of the damage is not exponential we have to solve equation (3.3.24) numerically in most cases. Of course we may also use the iteration cycle described in §2.

Literature:

1. G. de Leve Generalized Markovian Decision Processes,
Mathematical Centre Tract No. 3 and 4.