STICHTING
MATHEMATISCH CENTRUM
2e BOERHAAVESTRAAT 49
AMSTERDAM

STATISTISCHE AFDELING

Report S 282 (VP 15)

Decision Rules for adjusting Markovian Processes

by

G. de Leve

August 1961

The use of Markov processes as a tool for solving stochastic ∞-stage decision problems is well known.

In many papers situations are discussed in which at given discrete points of time decisions have to be made.

In this paper, however, we will consider a more general problem in which decisions may be taken at any timepoint. In a mainly verbal exposition we will make some remarks about this problem. In an other publication I hope to give the proofs of the results mentioned here.

In an ∞-stage decision problem we meet a system, the states of which can be described with state variables or, what is the same, can be identified with a point in a so called state space. This space will be indicated by ψ .

In the problem we have in mind the system, subjected to a Markov process, takes a random walk through the state space ψ . In each time interval the system produces losses. These losses, which are additive, depend only on the states of the system during that time interval and are independent of the states of the system outside that interval.

It will be clear that the decisionmaker, who is in charge, wants to prevent or at least wants to make improbable some of the system's most expensive excursions through state space.

In this paper it will be assumed that there are states in state space ψ in which the decisionmaker can intervene. An intervention results in a transition of the system from the intervention state (i.e. the state at the moment of intervention) into a new state. The new state does not belong to the intervention set with probability 1. From this state the system continues its walk. Each intervention however produces losses too.

In order to let the argument be as general as possible we suppose that on the intervention point the decisionmaker cannot fix the transition completely. He can merely choose from a class of transition distributions. These distributions fix the probabilities of the transition from the intervention states into sets of new initial states.

The choice of a set of intervention states and of a transition distribution for each of these intervention states will be called a strategy.

Obviously the original Markov-process indicated by S_0, is no longer suitable for describing the behaviour of the system. The application of a strategy χ creates a new stochastic process in the state space ψ, which will be called the adjusted process $S(\chi)$.

For more than one reason it will be convenient to introduce a series of adjusted processes $S_n(\chi)$ which are generated as follows:

a) The first n times interventions in the original process take place according to the strategy χ.

b) After these interventions we let the process go on undisturbed.

From the "point of view" of the system the processes $S_n(\chi)$ are alternative. The realisations of the processes $S_n(\chi)$ are equal up to the $(n+1)^{st}$ intervention point. From that point on only one of both processes can exist.

With the aid of the series of processes $S_n(\chi)$ the Markov property of the adjusted process $S(\chi)$ can be proved under rather weak assumptions concerning the nature of the original process S_0 and that of the strategy applied.

If we pay attention to the states of the intervention set only, the adjusted process then induces a time discrete Markov process.

The next problem to be considered is: "How to compare alternative strategies". As usual we will use the expected loss as a criterion for optimality.

In calculating the expected loss in an oo-stage decision problem with given discrete time points of intervention, one generally first computes the expected losses in time intervals between the interventions and then adds them either weighted or unweighted. It happens very often that in more complex situations this procedure leads to rather difficult probabilistic problems.

In order to avoid these problems we consider a product space ψ^* of a denumberable set of spaces $\psi_{I,i}$ $(i=0,1,\ldots)$ and the space ψ_{II}, which are all congruent to the space ψ.

Let us construct mathematically a new process $S^*(\chi)$ in such a way that the state of a system, subjected to this process, can be identified with a point of ψ^*.

The process $S^*(\chi)$ can be defined if we impose among other properties the following:

1) The process $S^*(\chi)$ induces in the subspace $\psi_{I,0}$ of ψ^* an S_0 process.

2) The process $S^*(\chi)$ induces in the subspace $\psi_{I,i}$ of ψ^* an $S_i(\chi)$ process.

3) The process $S^*(\chi)$ induces in the subspace ψ_{II} of ψ^* an $S(\chi)$ process.

4) According to the definition of an $S_n(\chi)$ process the induced paths of the system in the subspaces $\psi_{I,n}$ and $\psi_{I,n+1}$ are congruent up to the $(n+1)^{st}$ intervention point.

5) From the $(n+1)^{st}$ intervention point the processes $S_n(\chi)$ and $S_{n+k}(\chi)$ $(k=1,\ldots)$ are completely independent.

The process $S^*(\chi)$ is a mathematical abstraction and has no physical interpretation.

Next we introduce a new set of states in the state space ψ called the stopping set and indicated by E. The purpose of this set will be explained below. The decisionmaker is more or less free in the choice of this set. It is only subjected to the condition that from each initial state the system will almost surely arrive in the stopping set after a finite period of time.

Let us consider a new series of adjusted processes $S_n(\chi;E)$ generated as follows:

a) The first n times interventions in the original process take place according to the strategy χ.

b) At the $(n+1)^{st}$ intervention point we do not intervene but let the process go on indisturbed until a state of the stopping set is reached, where the process will be stopped.

The process $S_0(\chi;E)$ arises if we wait till the moment the system reaches a point of the intervention set and stop the process after that point of time as soon as the system takes on a state of the stopping set.

In this way we obtain a series of processes $S_n(\chi;E)$ with the following properties:

The processes $S_n(\chi;E)$ and $S_m(\chi;E)$ $(m < n)$ are identical up to and including the $(m+1)^{st}$ intervention point. From that point onwards, before the intervention, the system in the $S_n(\chi;E)$ process goes on as a system subjected to an $S_{n-m}(\chi;E)$ process, while the system in an $S_m(\chi;E)$ process continues its walk through state space like a system subjected to an $S_0(\chi;E)$ process.

If in a finite time interval a finite number of interventions take place with probability 1 the adjusted process can be regarded as the "limit" of a series of processes $S_n(\chi;E)$ with $n \to \infty$.

Let us return to the process $S^*(\chi)$. If a realisation of this process is given then one can easily verify that the realisation of each of the processes $S_n(\chi;E)$ is also given. This means that with probability 1 for each realisation of the process $S(\chi)$ and for each t a value n_0 can be found such that the corresponding realisations of the processes $S(\chi)$ and $S_n(\chi;E)$ are identical in the period (0,t), if $n \geq n_0$.

Let us suppose that the loss connected with a realisation of a process $S_j(\chi;E)$ (j=0,1,...) can be established and let us indicate this loss by l_j. The loss l_n, connected with a realisation of an $S_n(\chi;E)$ process, satisfies the obvious relation:

$$l_n = l_0 + \sum_{j=1}^{n} (l_j - l_{j-1}) \ . \qquad (1)$$

If the realisation of the process $S^*(\chi)$ is not given then the difference $l_j - l_{j-1}$ is, under certain measurability conditions, a stochastic variable. In that case the expected loss of an $S_n(\chi;E)$ process satisfies the relation:

$$\mathcal{E}\,\underline{l}_n = \mathcal{E}\,\underline{l}_0 + \sum_{j=1}^{n} \mathcal{E}(\underline{l}_j - \underline{l}_{j-1}) \ . \qquad (2)$$

According to the definitions of the processes $S_j(\chi;E)$ and $S_{j-1}(\chi;E)$ and of the losses of their realisations, the difference

$$\underline{l}_j - \underline{l}_{j-1} \qquad (3)$$

in (1) does not depend on the states of the system before the j^{th} intervention point.

Furthermore it follows from the definitions of the processes $S_j(\chi;E)$ and $S_{j-1}(\chi;E)$ that from and including the j^{th} intervention point they go on like the processes $S_1(\chi;E)$ and $S_0(\chi;E)$ respectively. If the realisations of the processes $S_j(\chi;E)$ and $S_{j-1}(\chi;E)$ are given then the realisations of the corresponding processes $S_1(\chi;E)$ and $S_0(\chi;E)$ are also known. Consequently the difference in (3) can also be considered as the difference in loss corresponding to these last mentioned realisations.

If the j^{th} intervention state is indicated by I_j, let $c(I_j;\chi)$ be the difference in expected losses of the processes $S_1(\chi;E)$ and $S_0(\chi;E)$ with initial state I_j.

From this definition it follows that:

$$\sum_{j=1}^{n} \mathcal{E}(\underline{l}_j-\underline{l}_{j-1}) = \sum_{j=1}^{n} \mathcal{E}\{c(I_j;\chi)\mid I_0\} \tag{4}$$

where I_0 is the initial state of the process.

Let $c(I_0;\chi)$ be the expected loss to be incurred by the system in an $S_0(\chi;E)$ process with initial state I_0. Then the expected loss, $\mathcal{E}\,\underline{l}_n$, to be incurred in an $S_n(\chi;E)$ process, from now indicated by $\check{c}_n(I_0;\chi)$, will be equal to:

$$\check{c}_n(I_0;\chi) = \sum_{j=0}^{n} \mathcal{E}\{c(\underline{I}_j;\chi)\mid I_0\}. \tag{5}$$

Let us suppose that the duration of each given realisation of a process $S_j(\chi;E)$ can be established and let us indicate this duration by d_j then of course we have:

$$d_n = d_0 + \sum_{j=1}^{n} (d_j-d_{j-1}) \tag{6}$$

If the $j^{\underline{th}}$ intervention state is indicated by I_j, let $t(I_j;\chi)$ be the difference in expected durations of the processes $S_1(\chi;E)$ and $S_0(\chi;E)$ with initial state I_j.

In the same way as we derived the relation (5) we can prove the following result:

If I_0 is the initial state of the process and $t(I_0;\chi)$ is the expected duration of an $S_0(\chi;E)$ process with that initial state then the expected duration of an $S_n(\chi;E)$ process, indicated by $\check{t}_n(I_0;\chi)$, will be equal to:

$$\check{t}_n(I_0;\chi) = \sum_{j=0}^{n} \mathcal{E}\{t(\underline{I}_j;\chi)\mid I_0\}. \tag{7}$$

Let us suppose that it is possible to put out to contract the charge of the system and suppose we have to pay in those circumstances a premium of α per unit of time. It will be cheaper to handle the system, subjected to an $S_n(\chi;E)$ process, ourselves if we have:

$$\sum_{j=0}^{n} \mathcal{E}\{c(\underline{I}_j;\chi)\mid I_0\} \leq \alpha \sum_{j=0}^{n} \mathcal{E}\{t(\underline{I}_j;\chi)\mid I_0\} \tag{8}$$

or

$$\frac{\frac{1}{n}\sum_{j=0}^{n}\mathcal{E}\{c(\underline{I}_j;\chi)\mid I_0\}}{\frac{1}{n}\sum_{j=0}^{n}\mathcal{E}\{t(\underline{I}_j;\chi)\mid I_0\}} \leq \alpha. \tag{9}$$

We have already stated that under rather weak assumptions the adjusted process induces a time discrete Markov process in the intervention set. If the functions $c(I;\chi)$ and $t(I;\chi)$ are defined for each I belonging to the intervention set, then we can restrict our attention to the time discrete Markov process mentioned above. Whether or not this process has a stationary absolute probability distribution depends on certain hypotheses concerning the structure of the process, e.g. Doeblin's hypothesis. [1] If with respect to this probability distribution the expected values of the functions

$$|c(\underline{I};\chi)| \text{ and } |t(\underline{I};\chi)|$$

exist then the expressions:

$$\frac{1}{n}\sum_{j=0}^{n} c(\underline{I}_j;\chi) \text{ and } \frac{1}{n}\sum_{j=0}^{n} t(\underline{I}_j;\chi) \tag{10}$$

converge with probability 1 for n tending to infinity [2].

Both limits depend on the ergodic set, in which the system "jumps".

If I_0 belongs to an ergodic set let the function $\bar{c}(I_0;\chi)$ and $\bar{t}(I_0;\chi)$ be the expected values of $c(\underline{I};\chi)$ and $t(\underline{I};\chi)$ respectively, with respect to the stationary absolute probability distribution corresponding to that ergodic set.

It can be proved that the following relations are true with probability 1: [2]

$$\lim_{n\to\infty} \frac{1}{n}\sum_{j=0}^{n} c(\underline{I}_j;\chi) = \bar{c}(I_0;\chi) \tag{11}$$

$$\lim_{n\to\infty} \frac{1}{n}\sum_{j=0}^{n} t(\underline{I}_j;\chi) = \bar{t}(I_0;\chi) \tag{12}$$

Let us suppose that there are k different ergodic sets and let I_{0h} $(h=1,\dots k)$ be states in these sets. If I_0 is a transient state and p_h is the probability of entering the $h^{\underline{th}}$ ergodic set from this state, then the limits of

$$\frac{1}{n}\sum_{j=0}^{n} c(\underline{I}_j;\chi) \text{ and } \frac{1}{n}\sum_{j=0}^{n} t(\underline{I}_j;\chi)$$

are stochastic quantities, which take on the values

$$\bar{c}(I_{0h};\chi) \text{ and } \bar{t}(I_{0h};\chi)$$

respectively with probability p_h.

Let p_h be the probability of entering the $h^{\underline{th}}$ ergodic set from a state $I_0 \in \psi_0$, then we have:

---

1) cf. J.L. Doob, Stochastic processes, p.192.
2) cf. J.L. Doob, Stochastic processes, p.220.

$$\mathcal{E}\Big[\lim_{n\to\infty}\frac{1}{n}\sum_{j=0}^{n}c(\underline{I}_j;\chi)\Big|I_0\Big] = \sum_{h=1}^{k}p_h\,\overline{c}(I_{0h};\chi) \tag{13}$$

$$\mathcal{E}\Big[\lim_{n\to\infty}\frac{1}{n}\sum_{j=0}^{n}t(\underline{I}_j;\chi)\Big|I_0\Big] = \sum_{h=1}^{k}p_h\,\overline{t}(I_{0h};\chi)\;. \tag{14}$$

Let us assume that I_0 belongs to an ergodic set and that the following inequalities are valid

$$\mathcal{E}\left\{\,|\,c(\underline{I};\chi)|^{2+\delta}\right\}<\infty \tag{15}$$

$$\mathcal{E}\left\{\,|\,t(\underline{I};\chi)|^{2+\delta}\right\}<\infty \tag{16}$$

for some $\delta > 0$.

The expected values in (15) and (16) are taken with respect to the stationary absolute probability distribution corresponding to the ergodic set of I_0.

It can be proved that [3)]

$$\lim_{n\to\infty}\frac{1}{\sqrt{n}}\sum_{m=0}^{n}\left[c(\underline{I}_m;\chi)-\mathcal{E}\,c(\underline{I}_m;\chi)\right] \tag{17}$$

is, for any initial distribution of $\underline{I}_0$ (thus also for $\underline{I}_0=I_0$) normally distributed with zero mean and finite variance.

From (17) it follows:

$$\mathcal{E}\left\{\lim_{n\to\infty}\frac{1}{n}\sum_{m=0}^{n}\left[c(\underline{I}_m;\chi)-\mathcal{E}\,c(\underline{I}_m;\chi)\right]\right\} = 0 \tag{18}$$

or

$$\mathcal{E}\Big[\lim_{n\to\infty}\frac{1}{n}\sum_{m=0}^{n}c(\underline{I}_m;\chi)\Big] = \lim_{n\to\infty}\frac{1}{n}\sum_{m=0}^{n}\mathcal{E}\,c(\underline{I}_m;\chi)\;. \tag{19}$$

In the same way we can prove:

$$\mathcal{E}\Big[\lim_{n\to\infty}\frac{1}{n}\sum_{m=0}^{n}t(\underline{I}_m;\chi)\Big] = \lim_{n\to\infty}\frac{1}{n}\sum_{m=0}^{n}\mathcal{E}\,t(\underline{I}_m;\chi)\;. \tag{20}$$

If I_0 is a transient state and if for each of the ergodic sets the inequalities (15) and (16) are satisfied then it can be proved that the relations (19) and (20) remain true.

Up to now we have made a number of assumptions concerning the nature of the process, the strategy applied and some special functions.

Strategies which do satisfy all these assumptions will be named admissible strategies.

Let us return to the inequality (9) and let us use the results expressed in the relations (13), (14), (19) and (20).

If χ is an admissible strategy and n tends to infinity 1, then

----

3) cf. J.L. Doob, Stochastic processes, p.228 and p.232.

the inequality (9) changes into:

$$\frac{\sum_{h=1}^{k} p_h \, \bar{c}(I_{0h};\chi)}{\sum_{h=1}^{k} p \;\; \bar{t}(I_{0h};\chi)} \leqslant \alpha \quad . \tag{21}$$

If the inequality (21) is true then it will be cheaper to handle the system, subjected to an $S(\chi)$ process, ourselves.

Let us suppose that, if we can show that "help yourself" is cheaper the substitute is willing to reduce the premium. If the reduced premium is equal to the left hand side of the inequality (21), then it will be clear that for an $S(\chi)$ process the admissible strategy, which minimizes the left hand side of (21) is the optimal one.

So we have derived for admissible strategies the criterion for optimality. The criterion function is given by:

$$\lambda(\chi;I_0) = \frac{\sum_{h=1}^{k} p_h \, \bar{c}(I_{0h};\chi)}{\sum_{h=1}^{k} p_h \, \bar{t}(I_{0h};\chi)} \quad . \tag{22}$$

In some simple problems the optimal strategy can be obtained directly from the criterion function. But, generally, the function $\lambda(\chi;I_0)$ does not have a simple explicit form.

In the remainder of this paper we should like to point out the fact that in this more general approach of stochastic ∞-stage decision problem a mathematical iteration procedure, a so called "policy improvement method" can also be developed.

In order to illustrate this we will introduce sequences of processes $S_{mn}(\chi_2;\chi_1;E)$ with $m \geqslant 1$ and $n \geqslant 0$. These processes can be generated as follows:

The first m interventions in the original process will take place in accordance with the strategy χ_2. After the m^{th} intervention the strategy χ_1 will be applied n times in succession.

Before the process will be stopped the system has to take on at least one other state in the intervention set of the strategy χ_1. From this definition of the $S_{mn}(\chi_2;\chi_1;E)$ process it follows that the $S_m(\chi_2;E)$ process may differ from the $S_{m0}(\chi_2;\chi_1;E)$ process. The expected value of the loss incurred in a $S_{mn}(\chi_2;\chi_1;E)$ process will be indicated by $\check{c}_{mn}(I_0;\chi_2;\chi_1)$, where I_0 is the initial state of the process.

If $\underline{I}_m^*$ is the state just <u>after</u> the $m^{\underline{th}}$ intervention then the following relation is easily to verify:

$$\check{c}_{mn}(I_0;\chi_2;\chi_1) =$$

$$= \check{c}_m(I_0;\chi_2)+ \mathcal{E}\{\check{c}_n(\underline{I}_m^*;\chi_1)|I_0\} - \mathcal{E}\{\check{c}_0(\underline{I}_m^*;\chi_2)|I_0\}. \qquad (23)$$

Let the expected duration of an $S_{mn}(\chi_1;\chi_2)$ process with initial state I_0 be given by $\check{t}_{mn}(I_0;\chi_2;\chi_1)$, then the following relation can be proved:

$$\check{t}_{mn}(I_0;\chi_2;\chi_1) = \check{t}_m(I_0;\chi_2)+\mathcal{E}\{\check{t}_n(\underline{I}_m^*;\chi_1)|I_0\} +$$

$$-\mathcal{E}\{\check{t}_0(\underline{I}_m^*;\chi_2)|I_0\}. \qquad (24)$$

Let us suppose we have a class of admissible strategies J with the following properties:

a) The class J contains also the strategy that comes into being if we have a series of strategies $\chi_j \in J$ with corresponding disjunct sets M_j of a certain Borel field, and if we apply the strategy χ_j always but only in M_j.

b) If $\chi \in J$ creates an $S_\infty(\chi)$ process in ψ with k ergodic sets, then corresponding to each of these sets there is a strategy $\chi^i (i=1,\dots k)$ conform to χ in that set and without any other ergodic set.

c) If $\chi_1 \in J$ has only one ergodic set then the strategy χ, that minimizes

$$H(I_0;\chi;\chi_1) =$$

$$= \lim_{n\to\infty}\{\check{c}_{1n}(I_0;\chi;\chi_1) - \lambda(\chi_1;I_0)\check{t}_{1n}(I_0;\chi;\chi_1) - \check{c}_n(I_0;\chi_1) +$$

$$+ \lambda(\chi_1;I_0)\check{t}_n(I_0;\chi_1)\} \qquad (25)$$

uniformly in I_0, belongs also to J.

Let us start then the "policy improvement" procedure with an arbitrary strategy $\chi_1 \in J$. If we apply the strategy χ_1 the adjusted process $S_\infty(\chi_1)$ comes into being. There are now two possibilities:

a) The adjusted process has only one ergodic set.

b) The adjusted process has more than one ergodic set.

If the adjusted process has more than one ergodic set, say k sets, let $\lambda_j(\chi_1)$ $(j=1,\dots k)$ be the corresponding values of the criterion function. If $\lambda_j(\chi_1)$ is minimal for $j=k_0$ then the transfer of the system to the $k_1^{\underline{th}}$ ergodic set has preference.

According to the property b) of the class J there is at least one strategy χ_1' that does not differ from χ_1 inside the $k_0^{\underline{th}}$ ergodic set and that has no other ergodic sets.

If the adjusted process has only one ergodic set we choose χ_1' equal to χ_1.

Our first step in the "policy improvement" procedure is to replace χ_1 by χ_1'.

Next we look for the strategy, indicated by χ_2, that minimizes:

$$H_0(I_0; \chi_2; \chi_1') =$$

$$\lim_{n \to \infty} \{ \check{c}_{1n}(I_0; \chi_2; \chi_1') - \lambda(\chi_1')\check{t}_{1n}(I_0; \chi_2; \chi_1') + \\ -\check{c}_n(I_0; \chi_1') + \lambda(\chi_1')\check{t}_n(I_0; \chi_1')\} \qquad (26)$$

uniformly in I_0.

According to the property c) of the class J the strategy χ_2 belongs to J.

By induction we can prove that the following inequality holds uniformly in I_0:

$$\check{c}_m(I_0; \chi_2) - \lambda(\chi_1')\check{t}_m(I_0; \chi_2) + \\ - \mathcal{E}\{\check{c}_0(\underline{I}_m^*; \chi_2) | I_0\} + \lambda(\chi_1')\mathcal{E}\{\check{t}_0(\underline{I}_m^*; \chi_2) | I_0\} + \\ + \lim_{n \to \infty} \mathcal{E}\{[\check{c}_{n-m}(\underline{I}_m^*; \chi_1') - \lambda(\chi_1')\check{t}_{n-m}(\underline{I}_m^*; \chi_1')] | I_0\} + \\ - \lim_{n \to \infty} \{\check{c}_n(I_0; \chi_1') - \lambda(\chi_1')\check{t}_n(I_0; \chi_1')\} \, . \leq 0 \, . \qquad (27)$$

If we pay only attention to the states just after an intervention it can be proved that under certain conditions, the adjusted process $S_{\infty}(\chi_2)$ induces a new time discrete Markov process with states I^* in state space.

Since the inequality mentioned above is valid for all I_0, we may assume that I_0 is distributed like one of the stationary absolute probability distributions of the new time discrete Markov process. If we take the expected value of the left hand side of (27) we will find:

$$\mathcal{E}[\check{c}_m(\underline{I}_0; \chi_2) - \lambda(\chi_1')\check{t}_m(\underline{I}_0; \chi_2)] + \\ - \mathcal{E}[\check{c}_0(\underline{I}_m^*; \chi_2) - \lambda(\chi_1')\check{t}_0(\underline{I}_m^*; \chi_2)] \leq 0. \qquad (28)$$

From the conditions imposed it follows that for $m \to \infty$ this inequality is, with probability 1, equivalent to:

$$\lim_{m \to \infty} \frac{\mathcal{E}\check{c}_m(\underline{I}_0; \chi_2)}{\mathcal{E}\check{t}_m(\underline{I}_0; \chi_2)} \leq \lambda(\chi_1') \qquad (29)$$

and consequently for each of the ergodic sets of $S_\infty(\chi_2)$:

$$\lambda(\chi_2; I_0) \leq \lambda(\chi_1'). \qquad (30)$$

This result implies that the strategy χ_2 has to be preferred to the strategy χ_1'.

An iteration method will be obtained if we restart the whole procedure with the strategy χ_2 instead of χ_1 and carry on in this way.

The iteration procedure thus produces a sequence χ_n'. For each n the inequality

$$0 \leq \lambda(\chi_n') \leq \lambda(\chi_{n-1}') \qquad (31)$$

holds.

Consequently the values of the criterion function corresponding to the sequence χ_n' converge to a value, to be indicated by λ_0.

We still have to prove that λ_0 is the value of the criterion function for the optimal strategy of J.

Suppose that the class J of admissible strategies has the additional property:

d) For each sequence of strategies χ_k' obtained from a "policy improvement" procedure the limit (in k) of

$$\lim_{n \to \infty} \{ \check{c}_{1n}(I_0; \chi_{k+1}'; \chi_k') - \lambda(\chi_k')\check{t}_{1n}(I_0; \chi_{k+1}'; \chi_k') \} \qquad (32)$$

and of

$$\lim_{n \to \infty} \{ \check{c}_{n+1}(I_0; \chi_k') - \lambda(\chi_k')\check{t}_{n+1}(I_0; \chi_k') \} \qquad (33)$$

converge uniformly in I_0 and are equal.

From the definition of χ_k' it follows, that for each k, for each strategy $\chi \in J$ and for each initial state I_0:

$$H(I_0; \chi; \chi_k') - H(I_0; \chi_{k+1}'; \chi_k') =$$

$$= \lim_{n \to \infty} \{ \check{c}_{1n}(I_0; \chi; \chi_k') - \lambda(\chi_k')\check{t}_{1n}(I_0; \chi; \chi_k') +$$

$$- \check{c}_{1n}(I_0; \chi_{k+1}'; \chi_k') + \lambda(\chi_k')\check{t}_{1n}(I_0; \chi_{k+1}'; \chi_k') \} \geq 0. \qquad (34)$$

If χ is the optimal strategy then from property d) it follows that for each $\varepsilon > 0$ a value k_0 can be found such that for $k > k_0$ and for each I_0 we have:

$$\lim_{n \to \infty} \{ \check{c}_{1n}(I_0; \chi; \chi'_k) - \lambda(\chi'_k)\check{t}_{1n}(I_0; \chi; \chi'_k) + \\ - \check{c}_{n+1}(I_0; \chi'_k) + \lambda(\chi'_k)\check{t}_{n+1}(I_0; \chi'_k)\} \geq -\varepsilon \quad . \qquad (35)$$

From this inequality we can deduce for each ergodic set of the $S_\infty(\chi)$ process:

$$\lambda(\chi; I_0) \geq \lambda_0 \qquad (36)$$

and consequently

$$\lambda(\chi; I_0) = \lambda_0 \quad . \qquad (37)$$

Postscript.

In this paper we have more than once used the expression "under certain conditions". The investigation with respect to these conditions is not yet fully completed.

The purpose of this paper is merely to outline a new method for solving some stochastic ∞-stage decision problems. In a number of concrete practical problems this method has already proved to be very useful.

Résumé

Dans cet article on considère des problèmes de décision Markoviennes, dans lesquelles on a la liberté de choisir sans restrictions les moments de décision.

La système considérée est soutenue d'un processus naturel de Markov. Si l'on prend une décision selon une politique fixée on peut troubler le processus naturel et par conséquence un processus nouvel s'élève. Sous des conditions spéciales imposées au processus naturel et à la politique appliquée le processus nouveau est de nouveau un processus de Markov.

Un critérium pour comparer les politiques est dérivé et une méthode itérative pour estimer la politique optimale est discutée.