STICHTING
MATHEMATISCH CENTRUM
2e BOERHAAVESTRAAT 49
AMSTERDAM

STATISTISCHE AFDELING

Rapport S 291

Meerstapsbeslissingsproblemen

door

G. de Leve

Lezing voor de Studiekring 's-Gravenhage van de Economische Sectie van de Vereniging voor Statistiek

November 1961

Bij het vermenigvuldigen van dit rapport is gebruik gemaakt van enkele stencils behorende bij rapport S 265.
De lezer zal hiervan geen hinder ondervinden, indien hij toevoegt aan blz. 11:

Uit (11) en (12) volgt:

$$q_{ij} = \sum_{k=1}^{n} q_{ik} p_{kj} \quad . \tag{16}$$

Indien de limiet

$$q_{ij} = \lim_{n \to \infty} \frac{1}{n} \sum_{k=1}^{n} p_{ij}^{(k)}$$

onafhankelijk is van de uitgangstoestand i dan gaat (16) over in

$$q_{j} = \sum_{k=1}^{n} q_{ik} p_{kj} \quad . \tag{17}$$

Voor de verwijzing op blz. 30 regel 16: (vgl. 6.8) leze men (vgl. 17).

Het feit dat er situaties bestaan waarin mensen beslissingen nemen zal ongetwijfeld niemand zijn ontgaan.

Evenzo mag bekend worden verondersteld dat in deze situaties genomen beslissingen verregaande gevolgen kunnen hebben.

In dit rapport willen wij eens nagaan in hoeverre wiskundige methoden een bijdrage kunnen leveren tot beter en sneller beslissen.

Steeds wanneer men wiskunde wil toepassen dan blijkt, dat slechts dan successen te verwachten zijn als alle facetten van het gestelde probleem zich laten vertalen in een wiskundige formulering. Hierbij dient echter te worden opgemerkt dat het bestaan van een mathematische formulering nog geen oplossing van het probleem garandeert.

Er kan slechts dan sprake zijn van een beslissingsprobleem als:

a) er meerdere beslissingen mogelijk zijn.

b) op grond van één of ander beginsel de mogelijke beslissingen niet gelijkwaardig zijn.

Uit de tekst van menig beslissingsprobleem blijkt, dat ook de toestand waarin een beslissing genomen wordt van invloed is op de keuze. Spreekt men echter van toestand dan moet ook iets bestaan waarop deze toestand betrekking heeft. In de wiskundige formulering spreekt men van de toestand van het systeem. Het systeem is veelal het object van onze beschouwingen. Zo kan bij een wachttijdprobleem het systeem gevormd worden door de rij wachtenden voor een loket. Bij een voorraadprobleem zal met het systeem meestal de voorraad worden bedoeld.

Aangenomen wordt dat het systeem zich in verschillende toestanden kan bevinden en dat deze toestanden beschreven kunnen worden door een aantal kwantitatieve grootheden. Een toestand wordt dus gegeven door een rij van getallen, de waarde van de kwantitatieve grootheden voor deze toestand.

Een dergelijke rij van getallen noemt men een <u>toestandsvector</u>. Indien wij een voorraad moeten beheren van drie verschillende artikelen dan is het systeem dus een voorraad. De toestand van het systeem, zo zullen wij aannemen, kan worden vastgelegd door drie getallen, die ieder de omvang van de voorraad van één artikel aangeven. Stel dat de toestanden van een systeem beschreven kunnen worden met behulp van m kwantitatieve grootheden, dan kan iedere toestand geidentificeerd worden met een punt in een door m orthogonale coördinaatassen opgespannen ruimte of met de toestandsvector, die de verbinding van dat punt met de oorsprong van het assenstelsel is. Op de assen van het assenstelsel worden de eerdergenoemde kwantitatieve grootheden uitgezet. Deze ruimte zullen wij de toestandsruimte noemen.

De mathematische formulering van het probleem beperkt zich uiteraard niet tot de beschrijving van de toestand van het systeem. We zullen aannemen, dat ook iedere mogelijke beslissing vastgelegd kan worden met behulp van een aantal kwantitatieve grootheden.

In het hierboven genoemde voorraadprobleem kan iedere beslissing worden vastgelegd door drie getallen. Deze getallen geven de ordergrootten aan van de drie artikelen. Voor het geval men de beslissingen kan aanduiden met behulp van r kwantitatieve grootheden, correspondeert met iedere beslissing een punt in een r dimensionale ruimte. De verbindingslijn van dat punt met de oorsprong zullen wij de beslissingsvector noemen, terwijl de ruimte zelf met beslissingsruimte wordt aangeduid.

Wij hebben reeds geconstateerd dat op grond van één of ander beginsel alle beslissingen niet gelijkwaardig zijn. Men zou ook kunnen zeggen, dat ten gevolge van een beslissing steeds sprake is van een opbrengst. [1] De omvang van deze opbrengst hangt af van de genomen beslissing en de toestand waarin deze beslissing is genomen.

Indien men de opbrengst aangeeft met y en de toestands- en beslissingsvector met x resp. d dan geldt

$$y = h(x,d) \qquad (1)$$

waarbij h(x,y) het ve band aangeeft tussen opbrengst enerzijds en toestand en beslissing anderzijds.

Het is niet ongebruikelijk om drie verschillende klassen van beslissingsproblemen te onderscheiden.

In de eerste klasse bevinden zich de zgn. één-stap-beslissingsproblemen. In een één-stap-beslissingsprobleem behoeft men slechts op één enkel tijdstip een beslissing te nemen.

In een N-stapsbeslissingsprobleem heeft men een rij van N tijdstippen, waarop achtereenvolgens een beslissing genomen dient te worden, terwijl in een ∞-stapsbeslissingsprobleem de lengte van

---

1) Er zijn problemen, waarbij de opbrengst stochastisch is. In die gevallen beschouwt men de verwachting van de opbrengst.

de rij onbeperkt is.

Zodra in een N-stapsbeslissingsprobleem de eerste beslissing genomen is gaat het N-stapsbeslissingsprobleem over in een (N-1) staps beslissingsprobleem.

Het ∞-staps beslissingsprobleem blijft na de eerste beslissing een ∞-staps beslissingsprobleem.

Deze beide opmerkingen zijn triviaal maar desondanks geven zij aan hoe oplossingen van de problemen uit beide groepen kunnen worden verkregen. Wij zullen dit aanstonds nader toelichten.

Indien wij aannemen dat in het begrip toestand alle relevante en beschikbare informatie over het systeem nu en in de toekomst is verwerkt dan zal de toestand van het systeem op het volgende tijdstip van beslissen sterk worden beinvloed door de toestand op het huidige tijdstip van beslissen en door de beslissing zelf, echter niet door de toestanden en beslissingen op daaraan voorafgaande tijdstippen.

Bij een definiet proces kan men nagaan wat de gevolgen zijn van een beslissing in een bepaalde toestand. Bij een proces met stochastische eigenschappen kan men niet met zekerheid voorspellen wat de gevolgen zijn van een beslissing in één of andere toestand.

Een voorbeeld van een meervoudig beslissingsprobleem,dat wiskundig geformuleerd kan worden is het schaakspel. Beperken wij onze aandacht tot de verrichtingen van één speler dan constateren wij dat deze speler in de loop van de tijd een reeks van beslissingen neemt. Aangezien de beslissingen sterk worden beïnvloed door de spelsituaties kan de speler voor het begin van het spel niet al zijn zetten kiezen.

Laten wij eens nagaan of het mogelijk is om de spelsituaties op een meer wiskundige wijze te onderscheiden. Hiertoe zouden wij b.v. een 32-dimensionaal orthogonaal assenstelsel kunnen invoeren en wel zo dat met iedere as een schaakstuk correspondeert. Aangezien het schaakbord bestaat uit 64 vakjes zullen wij op alle assen de getallen 1 t/m 64 uitzetten. Met ieder vakje van het bord correspondeert dan op elke as een punt. Met de oorsprong van elke as wordt aangeduid dat het desbetreffende stuk uit het veld is. Met behulp van deze punten kan men een 32-dimensionaal rooster construeren. Ieder roosterpunt geeft nu door zijn projecties op de assen een spelsituatie weer. Dit rooster vervult de rol van toestandsruimte. Op grond van de spelregels zijn slechts een beperkt aantal overgangen van roosterpunt tot roosterpunt mogelijk.

Indien men aanneemt dat zowel speler als tegenspeler geoefend zijn dan kunnen vier verzamelingen van roosterpunten worden aangegeven. In de eerste en tweede verzameling bevinden zich de situaties welke voor de speler respectievelijk tegenspeler noodlottig zijn. In de derde verzameling is het remise, terwijl de strijd nog open is in situaties van de laatste verzameling. Voor de beslissingsruimte kiezen wij wederom een 32-dimensionaal rooster, dat congruent is aan de toestandsruimte.

Geven wij de toestand van het systeem aan met een vector (verbinding oorsprong-roosterpunt) en wordt de beslissing eveneens vastgelegd met een vector dan kunnen wij de toestand x_k van het systeem juist voor de $k^{\underline{de}}$ zet van de speler geven door

$$x_k = x_{k-1} + d_{k-1} + d'_{k-1} \qquad (2)$$

waarbij d_{k-1} en d'_{k-1} de vectoren zijn behorende bij de $(k-1)^{ste}$ zet van de speler resp. tegenspeler.

Indien men alle stukken van een waarde voorziet en wel zo dat stukken van één soortgelijke waarden bezitten dan beginnen beide spelers met evenveel punten. Men spreekt nu af dat iedere speler bij zijn score mag optellen de waarden van de door hem geslagen stukken. De punten van de stukken die verloren zijn gegaan dienen echter te worden afgetrokken. Met ieder roosterpunt x correspondeert nu een getal $S(x)$, de score van de speler.

De opbrengst van de $(k-1)^{ste}$ zet $h(x_{k-1},d_{k-1})$ wordt nu gegeven door:

$$h(x_{k-1},d_{k-1}) = S(x_{k-1}+d_{k-1}+d'_{k-1}) - S(x_{k-1}). \qquad (3)$$

In éénstaps beslissingsproblemen zullen wij een beslissing optimaal noemen als voor deze beslissing de opbrengst maximaal is.

De vraag is nu hoe kunnen wij deze definitie uitbreiden tot n-staps beslissingsproblemen?

Van de optimale beslissing in een n-staps beslissingsprobleem wordt verwacht, dat zij de opbrengst van alle n-beslissingen tezamen maximaliseert. Aangezien deze beslissing mede de toestand bepaalt op het volgende beslissingstijdstip en dientengevolge de opbrengst zal zij in het algemeen niet gelijk zijn aan de beslissing welke alleen de directe opbrengst maximaliseert. Om het effect van een beslissing op toekomstige opbrengsten te kunnen nagaan, dient men, omdat de opbrengsten hierdoor mede worden bepaald, de overige beslissingen te kennen.

Men kan dus alleen dan de optimale $(k-1)^{ste}$ zet van het schaakspel bepalen als men weet voor iedere $(k-1)^{ste}$ zet welke de volgende zetten zullen zijn van speler en tegenspeler. Het is duidelijk, dat wij voor wat het schaakspel betreft in een impasse zijn beland. Dit laatste is dan ook niet te verwonderen. Wij zullen ons tot eenvoudiger problemen moeten beperken.

In het hierna volgende bezitten de problemen steeds één van de volgende eigenschappen.

1. De toestand op het $k^{\underline{de}}$ beslissingstijdstip wordt volledig bepaald door de toestand op het $(k-1)^{ste}$ beslissingstijdstip en de daarop genomen beslissing.

2. De kansverdeling van de mogelijke toestanden op het $k^{\underline{de}}$ beslissingstijdstip wordt volledig bepaald door de toestand op het $(k-1)^{ste}$ beslissingstijdstip en de daarop genomen beslissing.

Het schaakprobleem voldoet noch aan de eerste noch aan de tweede eigenschap en kan derhalve niet met de hier te behandelen methoden worden aangepakt.

Ondanks het feit dat wij ons zullen beperken tot minder ingewikkelde problemen, de optimale eerste beslissing van een n-staps beslissingsprobleem kan men nog steeds niet bepalen zonder de latere beslissingen te kennen.

<u>N-staps dynamisch programmeringsprobleem</u>

Aan deze moeilijkheid kan op de volgende wijze het hoofd worden geboden. Men begint nl. bij een n-staps d.p. probleem niet met de bepaling van de <u>eerste</u> optimale beslissing maar met de laatste. Hiervoor dient men de toestand te kennen op het bijbehorende tijdstip. Aangezien men die toestand niet kent lost men het probleem op voor <u>alle</u> mogelijke toestanden. De maximale opbrengst is derhalve een functie van de toestand op het laatste beslissingstijdstip.

Deze toestanden zelf of de kansverdeling behorende bij deze toestanden worden volledig bepaald door de toestand van het systeem en de beslissing op het voorgaande moment van beslissen. De maximale opbrengst kan dus beschouwd worden als een functie alleen van die grootheden. Indien men een nieuwe opbrengstfunctie invoert, die gelijk is aan de oude vermeerderd met de maximum opbrengst van de volgende beslissing, dan kan deze nieuwe opbrengstfunctie gemaximaliseerd worden, waardoor de optimale beslissing voor het op één na laatste moment van beslissen wordt verkregen.

Ook nu geldt dat de toestand van het systeem op het beslissingstijdstip niet bekend is zodat het maximum een functie is van deze toestand.

Resumerend kunnen wij vaststellen, dat wij eerst 1-staps problemen oplossen voor alle mogelijke uitgangstoestanden van het systeem en daarna met behulp van deze oplossingen 2-staps problemen oplossen voor alle mogelijke begintoestanden.

Indien men op deze wijze voortgaat, wordt de oplossing van het oorspronkelijke gestelde vraagstuk verkregen.

Bij het oplossen van een d.p. probleem wordt het oorspronkelijke probleem dus vervangen door een reeks van in eenvoud afnemende problemen, welke worden opgelost voor alle mogelijke begintoestanden. Indien wij de toestand van het systeem en de beslissing op het $m^{\underline{de}}$ beslissingstijdstip resp. aangeven met x_m en d_m en als wij verder schrijven voor de bijbehorende opbrengst $h(x_m, d_m)$ en voor de toekomstige of verwachte toekomstige opbrengsten $H_{n-m}(x_m, d_m)$ bij optimale keuze van d_{m+1}, d_{m+2} enz. dan geldt voor een n-staps d.p. probleem:

$$H_{n-m}(x_m, d_m) = \mathcal{E} \max_{d_{m+1}} \left[h(\underline{x}_{m+1}, d_{m+1}) + H_{n-m-1}(\underline{x}_{m+1}, d_{m+1}) \right] \quad (m=n-1) \qquad (4)$$

waarbij

$$H_o(x_n, d_n) = 0 \,. \qquad (5)$$

Uit deze beschouwing volgt tevens, dat bij een n-staps d.p. probleem de eerste optimale beslissing d_1 alleen bepaald wordt door de toestand x_1. Men zou ook kunnen zeggen dat door de optimale keuze aan een toestandsvector in de toestandsruimte een beslissingsvector in de beslissingsruimte wordt toegevoegd. Deze toevoeging kan men ook als volgt uitdrukken:

$$d_1 = S_n(x_1). \qquad (6)$$

Bellman heeft door gebruik te maken van het abstracte begrip Functieruimte de zgn. <u>strategieruimte</u> ingevoerd. Ieder punt uit deze ruimte is een <u>functie</u> met behulp waarvan aan iedere <u>toestandsvector</u> een beslissingsvector kan worden toegevoegd. Ook de functie $S_n(x)$ in (6) is een punt uit deze ruimte en wordt een strategie-functie of ook wel kortweg strategie genoemd.

Wij zullen nu tot slot een toepassing geven.

Een reisbureau heeft voor een periode van 6 jaar een hotel gepacht in een wintersportcentrum. Met een plaatselijke kolenhandelaar is een contract afgesloten, waarin wordt bepaald dat de kolenhandelaar elk jaar een vaste hoeveelheid brandstof zal leveren tegen een in het contract genoemd bedrag. Verder mag het reisbureau in geval van wanprestatie het contract éénzijdig opzeggen.

De kolenhandelaar verkoopt 3 soorten kolen en wel de volgende:

a) superkolen
b) kwaliteitskolen
c) huishoudkolen.

Wij zullen de bij de verkoop per jaar te maken winst aangeven met a_i. $(i=1,2,3)$ De kolensoorten zijn genummerd in de hierboven aangegeven volgorde. Voor de getallen a_i geldt de ongelijkheid:

$$a_1 < a_2 < a_3 \ .$$

De kans op opzegging van het contract na levering van de i^{de} kolensoort wordt aangeduid met p_i en is voor elk jaar gelijk. Voor de kansen p_i geldt de ongelijkheid:

$$p_1 < p_2 < p_3 \ .$$

De kolenhandelaar, een harde zakenman, vraagt zich af welke kolensoorten hij in de opéénvolgende jaren bij het hotel zal afleveren.

<u>Oplossing</u> (beknopt):

Het systeem in dit probleem is de kolentransactie. De toestand van het systeem wordt gegeven door

$x = 0$ contract opgezegd
$x = 1$ contract geldig.

De beslissingen worden gegeven door:

$d = 1$ superkolen
$d = 2$ kwaliteitskolen
$d = 3$ huishoudkolen.

De relatie (4) gaat over in:

$$H_{6-m}(0;d_m) = 0.$$
$$H_{6-m}(1;d_m) = (1-p_{d_m}) \max_{d_{m+1}} \left[a_{d_{m+1}} + H_{6-(m+1)}(1;d_{m+1}) \right] .$$

Stel:

$$a_1 = 435 \qquad\qquad p_1 = 0{,}2$$
$$a_2 = 790 \qquad \text{en} \qquad p_2 = 0{,}4$$
$$a_3 = 1050 \qquad\qquad p_3 = 0{,}6,$$

dan vinden wij:

$$H_1(0,d_5) = 0$$
$$H_1(1;1) = (1-0{,}2)\max_{d_6}\left[a_{d_6}\right] = 0{,}8 \times 1050 = 840$$
$$H_1(1;2) = (1-0{,}4)\max_{d_6}\left[a_{d_6}\right] = 0{,}6 \times 1050 = 630$$
$$H_1(1;3) = (1-0{,}6)\max_{d_6}\left[a_{d_6}\right] = 0{,}4 \times 1050 = 420$$
$$d_6 = 3$$

$$H_2(0;d_4) = 0$$
$$H_2(1;1) = 0{,}8 \times \max_{d_5}\left[a_{d_5}+H_1(1;d_5)\right] = 1176$$
$$H_2(1;2) = 0{,}6 \times \max_{d_5}\left[a_{d_5}+H_1(1;d_5)\right] = 882$$
$$H_2(1;3) = 0{,}4 \times \max_{d_5}\left[a_{d_5}+H_1(1;d_5)\right] = 588$$
$$d_5 = 3$$

$$H_3(0;d_3) = 0$$
$$H_3(1;1) = 0{,}8 \times \max_{d_4}\left[a_{d_4}+H_2(1;d_4)\right] = 1337{,}60$$
$$H_3(1;2) = 0{,}6 \times \max_{d_4}\left[a_{d_4}+H_2(1;d_4)\right] = 1003{,}20$$
$$H_3(1;3) = 0{,}4 \times \max_{d_4}\left[a_{d_4}+H_2(1;d_4)\right] = 668{,}80$$
$$d_4 = 2$$

$$H_4(0;d_2) = 0$$
$$H_4(1;1) = 0{,}8 \times \max_{d_3}\left[a_{d_3}+H_3(1;d_3)\right] = 1434{,}56$$
$$H_4(1;2) = 0{,}6 \times \max_{d_3}\left[a_{d_3}+H_3(1;d_3)\right] = 1075{,}92$$
$$H_4(1;3) = 0{,}4 \times \max_{d_3}\left[a_{d_3}+H_3(1;d_3)\right] = 717{,}38$$
$$d_3 = 2$$

$$H_5(0;d_1) = 0$$
$$H_5(1;1) = 0{,}8 \times \max_{d_2} [a_{d_2} + H_4(1;d_2)] = 1495{,}65$$
$$H_5(1;2) = 0{,}6 \times \max_{d_2} [a_{d_2} + H_4(1;d_2)] = 1121{,}74$$
$$H_5(1;3) = 0{,}4 \times \max_{d_2} [a_{d_2} + H_4(1;d_2)] = 747{,}82$$
$$d_2 = 1$$

De optimale eerste beslissing vindt men nu uit het max.van

$$\{ a_{d_1} + H_5(1;d_1) \}$$

of $\{ 1.930{,}65;\ 1.911{,}74;\ 1.797{,}82 \}$.
De optimale eerste beslissing is dus $d_1 = 1$.

∞-staps dynamisch programmeringsprobleem

Wij zullen ons nu bezighouden met situaties, waarin het aantal te nemen beslissingen onbegrensd is. Het is zonder meer duidelijk, dat de oplossing van deze problemen niet verkregen wordt door eerst de beslissing van de "laatste" stap te bepalen.

Stel dat de maximale opbrengst bij een n-staps d.p. probleem gegeven wordt door de functie $f_n(x)$, waarbij x de toestandsvector aanduidt op het eerste tijdstip van beslissen.

Uit de definitie van $H_i(x,d)$ volgt:

$$f_n(x) = \max_d \{ h(x,d) + H_{n-1}(x,d) \}. \qquad (7)$$

Indien uit de probleemstelling volgt:

a) dat de kansverdeling van de toestandsvector behorende bij het volgende beslissingstijdstip volledig bepaald wordt door de toestand op het huidige moment van beslissen en de beslissing zelf.

b) dat op het $i^{\underline{de}}$ beslistijdstip slechts rekening wordt gehouden met een fractie van de opbrengsten behorende bij toekomstige beslissingen en wel zo dat de bijdrage van de $j^{\underline{de}}$ beslissing ($j \geq i$) tot de totale opbrengst gelijk is aan:

$$\alpha^{j-i}\, h(x_j;d_j)\ . \qquad (8)$$

Schrijft men nu voor de toestandsvector en beslissingsvector behorend bij het $i^{\underline{de}}$ beslissingstijdstip, x_i resp. d_i, dan vinden wij voor de opbrengstfunctie $f_n(x)$:

$$f_n(x_1) = \max_{d_1} \left[h(x_1,d_1) + \alpha E \left\{ f_{n-1}(\underline{x}_2)/x_1;d_1 \right\} \right] . \qquad (9)$$

Beschouwen wij nu de rij van functies $f_n(x)$ dan zal, als deze rij voor iedere x convergeert tot een functie f(x), (9) overgaan in:

$$f(x_1) = \max_{d_1} \left[h(x_1,d_1) + \alpha E \left\{ f(\underline{x}_2)/x_1;d_1 \right\} \right] . \qquad (10)$$

Indien de opbrengstfuncties $f_n(x)$ behorende bij de overeenkomende n-staps d.p. problemen convergeren tot de functie f(x), dan wordt de oplossing van het ∞-staps d.p. probleem verkregen uit de oplossing van de corresponderende functionaalvergelijking. Het onderzoek naar de ondubbelzinnigheid en de existentie van de oplossingen van deze functionaalvergelijkingen vormt een wezenlijk deel van de dynamische programmering. Aan dit belangrijke onderdeel van de theorie der dynamische programmering zullen wij echter geen aandacht besteden.

Markovian decision processes

Wij beschouwen wederom een ∞-staps beslissingsprobleem, waarbij men op van te voren vastgestelde tijdstippen een beslissing kan nemen.

Indien alle relevante informatie over het systeem nu en in de toekomst verwerkt is in het begrip toestand dan zal men elke keer als het systeem op een beslissingstijdstip in een bepaalde toestand komt dezelfde beslissing nemen. Hieruit volgt dat men de optimale strategie kan vastleggen door aan ieder punt uit de toestandsruimte een punt uit de beslissingsruimte toe te voegen. Deze beslissing wordt slechts dan uitgevoerd als het systeem deze toestand aanneemt op een beslissingstijdstip.

Nu bestaat er een uitgebreide klasse van problemen, waarvan de toestandsveranderingen van het systeem beschreven kunnen worden door een Markov proces. Dit Markov proces hangt af van de keuze van de strategie. Voor verschillende strategieën krijgt men verschillende Markov processen. Ter vereenvoudiging van de discussie zullen wij aannemen, dat het systeem zich slechts in eindig veel toestanden i kan bevinden (i=1,...,n).

Stel dat de kans op een overgang van toestand i nu naar een toestand j op het volgende beslissingstijdstip gegeven wordt door $p_{ij}(x)$, als x de toegepaste strategie voorstelt. Indien $p_{ij}^{(k)}(x)$ de kans voorstelt op een overgang van i nu naar een toestand j op k beslissingstijdstippen verder dan geldt:

$$p_{ij}^{(k)}(x) = \sum_{l=1}^{n} p_{il}^{(k-1)}(x)\ p_{lj}(x) \quad . \qquad (11)$$

Nu kan men bewijzen, dat de uitdrukking

$$\lim_{n \to \infty} \frac{1}{n} \sum_{k=1}^{n} p_{ij}^{(k)}(x) = q_{ij}(x) \qquad (12)$$

bestaat. Men noemt de grootheden q_{ij} de invariante kansen van het Markov proces.

Stel dat $h(i,x)$ de opbrengst is, wanneer men in de toestand i een beslissing neemt volgens x. De verwachting van de totale opbrengst in m stappen wordt dan gegeven door:

$$\sum_{k=1}^{m} \sum_{j=1}^{n} p_{ij}^{(k)}(x)h(j;x) \quad . \qquad (13)$$

Indien men een strategie zoekt waarvoor de opbrengst voor de eerste m stappen maximaal is, dan moet voor deze strategie gelden dat (13) maximaal is. Het maximaliseren van de uitdrukking (13) is gelijkwaardig met het maximaliseren van

$$\frac{1}{m} \sum_{k=1}^{m} \sum_{j=1}^{n} p_{ij}^{(k)}(x)h(j;x) = \sum_{j=1}^{n} \left[\frac{1}{m} \sum_{k=1}^{m} p_{ij}^{(k)}(x) \right] h(j;x). \quad (14)$$

Indien $m \to \infty$ volgt uit (13) en (14) dat men als criterium moet kiezen voor de optimale strategie:

$$R(x) = \sum_{j=1}^{n} q_{ij}(x)h(j;x) \quad . \qquad (15)$$

Met behulp van (15) kan men dus twee alternatieve strategieën vergelijken.

Een verzekeringsprobleem

Een transportonderneming heeft bij een assurantiemaatschappij een schadeverzekering afgesloten. De voorwaarden, welke in de polis zijn beschreven, kunnen als volgt worden samengevat:

1) De verzekeringsmaatschappij kent 4 klassen van verzekerden. De transportonderneming betaalt in de eerste klasse een premie van ƒ 1.200,-. Voor de overige klassen bedraagt de premie resp. ƒ 900,-, ƒ 700,- en ƒ 600,-.
   Indien de transportonderneming gedurende een kalenderjaar geen schade heeft geclaimd, dan wordt zij het volgende jaar, tenzij zij reeds tot de hoogste klasse behoort, ingedeeld in de daaropvolgende

hcgere klasse.

Voor het geval zij reeds tot de hoogste klasse behoort, blijft zij tot die klasse behoren. Indien echter de transportonderneming een claim tot schadevergoeding heeft ingediend, dan wordt zij het volgende jaar gerekend te behoren tot de laagste klasse.

2) De transportonderneming behoeft slechts aan het eind van ieder jaar de totale schade op te geven.

Er wordt dan en slechts dan een schadevergoeding uitgekeerd als het schadebedrag groter is dan

a) ƒ 750,- voor de eerste klasse
b) ƒ 600,- voor de tweede klasse
c) ƒ 500,- voor de derde klasse
d) ƒ 450,- voor de vierde klasse

De laatstgenoemde bedragen zijn de z.g. eigen risico's.

3) De uit te keren schadevergoeding is gelijk aan het schadebedrag verminderd met het eigen risico.

Het is zonder meer duidelijk, dat de transportonderneming in de eerste klasse geen schade zal claimen welke kleiner is dan ƒ 1.050,-. Immers zij derft een premieverlaging voor het komende jaar, welke groter is dan de schadeuitkering. Ook voor de overige klassen kan men tot soortgelijke uitspraken geraken. Wellicht is het voor de eerste klasse nog voordeliger om de "claimgrens" hoger te kiezen dan ƒ 1.050,-.

In deze toepassing zullen wij trachten de optimale "claimgrenzen" α_i ($i=1,\ldots4$) voor de vier klassen te berekenen.

Eén en ander hangt uiteraard ook af van de kansverdeling van de totale schade per jaar. Indien de omvang van de schade wordt aangeduid met t, dan wordt de verdelingsfunctie $F(t)$ van t gegeven door de kansdichtheid $f(t)$:

$$f(t) = \frac{1}{1000}\, e^{-\frac{t}{1000}}$$

Oplossing:

Uit het gestelde volgt, dat de transportonderneming, gezien als systeem, zich in 4 toestanden kan bevinden. Deze toestanden zullen wij overeenkomstig het nummer van de klasse aangeven met de getallen 1 t/m 4.

Aangezien men in de eerste klasse geen schade zal claimen, welke kleiner is dan α_1 wordt de overgangswh p_{12} gegeven door $F(\alpha_1)$. Op

dezelfde wijze vindt men de overgangswhn p_{23} en p_{34}, welke resp. gegeven worden door $F(\alpha_2)$ en $F(\alpha_3)$.
De overgangswh p_{44} is gelijk aan $F(\alpha_4)$.
Indien de transportonderneming tot de eerste klasse van verzekerden behoort, zal zij het volgende jaar tot de eerste of tot de tweede klasse behoren. Dit betekent dat $p_{11} + p_{12} = 1$, zodat p_{11} gelijk is aan $1-F(\alpha_1)$.
Op dezelfde wijze vinden we dat de overgangswhn p_{21}, p_{31} en p_{41} resp. gelijk zijn aan $1-F(\alpha_2)$, $1-F(\alpha_3)$ en $1-F(\alpha_4)$.
De matrix der overgangswhn heeft dus de volgende gedaante:

$$\begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \end{array} \begin{array}{c} \begin{array}{cccc} \quad 1 \quad & \quad 2 \quad & \quad 3 \quad & \quad 4 \quad \end{array} \\ \begin{pmatrix} 1-F(\alpha_1) & F(\alpha_1) & 0 & 0 \\ 1-F(\alpha_2) & 0 & F(\alpha_2) & 0 \\ 1-F(\alpha_3) & 0 & 0 & F(\alpha_3) \\ 1-F(\alpha_4) & 0 & 0 & F(\alpha_4) \end{pmatrix} \end{array} \qquad (11.2)$$

De invariante kansen q_1 t/m q_4 moeten nu voldoen aan de volgende betrekkingen: (vgl. 6.8)

$$\begin{aligned} q_1 &= [1-F(\alpha_1)]q_1 + [1-F(\alpha_2)]q_2 + [1-F(\alpha_3)]q_3 + [1-F(\alpha_4)]q_4 \\ q_2 &= F(\alpha_1)q_1 \\ q_3 &= F(\alpha_2)q_2 \\ q_4 &= F(\alpha_3)q_3 + F(\alpha_4)q_4. \end{aligned} \qquad (11.3)$$

Hieruit volgt:

$$q_i = \frac{\prod_{j=1}^{i-1} F(\alpha_j)}{\sum_{h=1}^{3} \prod_{k=1}^{h-1} F(\alpha_k) + \frac{\prod_{k=1}^{3} F(\alpha_k)}{1-F(\alpha_4)}} \quad (i=1,2,3)$$

$$q_4 = \frac{\frac{\prod_{k=1}^{3} F(\alpha_k)}{1-F(\alpha_4)}}{\sum_{h=1}^{3} \prod_{k=1}^{h-1} F(\alpha_k) + \frac{\prod_{k=1}^{3} F(\alpha_k)}{1-F(\alpha_4)}} \qquad (11.4)$$

Als het systeem zich in de i^{de} toestand bevindt, dan worden de volgende kosten gemaakt:

1) de premie A_i

2) de schade $\underline{t}_i$ als $\underline{t}_i \leqslant \alpha_i$

de schade β_i als $\underline{t}_i > \alpha_i$, waarin β_i het eigen risico voorstelt.

De verwachting van de schade voor deze toestand bedraagt dus:

$$A_i + \int_0^{\alpha_i} tf(t)dt + \beta_i \int_{\alpha_i}^{\infty} f(t)dt \,. \tag{11.5}$$

Voor de verwachting van de schade voor een "heel ver in de toekomst liggend" jaar vinden wij dus:

$$H(\alpha_1, \ldots, \alpha_4) = \sum_{i=1}^{4} q_i \left\{ A_i + \int_0^{\alpha_i} tf(t)dt + \beta_i \int_{\alpha_i}^{\infty} f(t)dt \right\} \tag{11.6}$$

Een stel waarden α_i $(i=1,\ldots,4)$ kan men een strategie noemen en zoals in par. 9 is aangetoond, zijn die waarden van α_i optimaal, waarvoor de uitdrukking (11.6) minimaal wordt.

De optimale waarden van α_i moeten voldoen aan de relaties:

$$\frac{\partial_H (\alpha_1, \ldots, \alpha_4)}{\partial \alpha_i} = 0 \,. \qquad (i=1,2,3,4) \tag{11.7}$$

Deze voorwaarden zijn wel noodzakelijk, echter niet voldoende.
Aangezien q_j een functie is van α_i, is (11.7) gelijkwaardig met:

$$\sum_{j=1}^{4} \frac{\partial q_j}{\partial \alpha_i} \left\{ A_j + \int_0^{\alpha_j} tf(t)dt + \beta_j \int_{\alpha_j}^{\infty} f(t)dt \right\} + q_i(\alpha_i - \beta_i) f(\alpha_i) = 0. \tag{11.8}$$

Uit (11.4) volgt voor $\frac{\partial q_j}{\partial \alpha_i}$:

a) $$\frac{\partial q_j}{\partial \alpha_i} = q_j \sum_{h=1}^{i} q_h \frac{f(\alpha_i)}{F(\alpha_i)} \qquad \text{als} \quad j \geqslant i+1, \quad i \leqslant 3$$

b) $$\frac{\partial q_j}{\partial \alpha_i} = q_j \sum_{h=1}^{i} q_h \frac{f(\alpha_i)}{F(\alpha_i)} - q_j \frac{f(\alpha_i)}{F(\alpha_i)} \quad \text{als } j \leqslant i,\ i \leqslant 3$$

$$(11.9)$$

c) $$\frac{\partial q_j}{\partial q_4} = -q_j q_4 \frac{f(\alpha_4)}{1-F(\alpha_4)} \qquad \text{als} \quad j \leqslant 3$$

d) $$\frac{\partial q_4}{\partial \alpha_4} = q_4(1-q_4) \frac{f(\alpha_4)}{1-F(\alpha_4)}$$

Voor de relatie (11.8) mogen wij dus ook schrijven voor $i \leqslant 3$:

$$\frac{f(\alpha_i)}{F(\alpha_i)} \sum_{h=1}^{i} q_h \sum_{j=1}^{4} q_j \left\{ A_j + \int_0^{\alpha_j} tf(t)dt + \beta_j \int_{\alpha_j}^{\infty} f(t)dt \right\} +$$

$$- \frac{f(\alpha_i)}{F(\alpha_i)} \sum_{j=1}^{i} q_j \left\{ A_j + \int_0^{\alpha_j} tf(t)dt + \beta_j \int_{\alpha_j}^{\infty} f(t)dt \right\} +$$

$$+ q_i(\alpha_i - \beta_i) f(\alpha_i) = 0 \qquad (11.10)$$

ofwel

$$\frac{f(\alpha_i)}{F(\alpha_i)} \sum_{h=1}^{i} q_h \left[H(\alpha_1, \ldots, \alpha_4) - H_i(\alpha_1, \ldots, \alpha_4) + \frac{q_{i+1}}{\sum_{k=1}^{i} q_k} (\alpha_i - \beta_i) \right] = 0 \qquad (11.11)$$

waarbij

$$H_i(\alpha_1, \ldots, \alpha_4) = \sum_{j=1}^{i} \frac{q_j}{\sum_{h=1}^{i} q_h} \left\{ A_j + \int_0^{\alpha_j} tf(t)dt + \beta_j \int_{\alpha_j}^{\infty} f(t)dt \right\}. \qquad (11.12)$$

Uit (11.11) volgt:

a) $\dfrac{f(\alpha_i)}{F(\alpha_i)} = 0$

of

b) $\sum_{h=1}^{i} q_h = 0$ d.w.z. $q_h = 0 \quad h \leq i$ (11.13)

of

c) $H(\alpha_1, \ldots, \alpha_4) - H_i(\alpha_1, \ldots, \alpha_4) + \dfrac{q_i F(\alpha_i)}{\sum_{h=1}^{i} q_h} (\alpha_i - \beta_i) = 0.$

Uit (11.13a) en (11.1) volgt $\alpha_i = \infty$.
Deze waarden van α_i voldoen aan (11.7). Langs wiskundige weg kan men aantonen dat ze toch niet optimaal zijn, hetgeen ook op intuïtieve gronden voor de hand ligt; immers er wordt wel premie betaald, doch er worden nooit schaden geclaimd.
Uit (11.13b) en (11.4) volgt $F(\alpha_4)=1$, m.a.w. in de 4de klasse wordt geen schade geclaimd. Ook deze beslissing is niet de beste.
Na deze overwegingen blijft dus over de relatie (11.13c).
Voor i=4 gaat de relatie (11.8) over in:

$$-q_4 \frac{f(\alpha_4)}{1-F(\alpha_4)} \sum_{j=1}^{4} q_j \left\{ A_j + \int_0^{\alpha_j} tf(t)dt + \beta_j \int_{\alpha_j}^{\infty} f(t)dt \right\} +$$

$$+ q_4 f(\alpha_4)(\alpha_4 - \beta_4) + q_4 \frac{f(\alpha_4)}{1-F(\alpha_4)} \left\{ A_4 + \int_0^{\alpha_4} tf(t)dt + \beta_4 \int_{\alpha_4}^{\infty} f(t)dt \right\} = 0 \tag{11.14}$$

ofwel

$$\frac{q_4 f(\alpha_4)}{1-F(\alpha_4)} \Big\{ A_4 + \int_0^{\alpha_4} tf(t)dt + \beta_4 \int_{\alpha_4}^{\infty} f(t)dt - H(\alpha_1, \ldots, \alpha_4) +$$

$$+ (\alpha_4 - \beta_4)\left[1 - F(\alpha_4)\right] \Big\} = 0. \tag{11.15}$$

Uit (11.15) volgt:

a) $\dfrac{f(\alpha_4)}{1-F(\alpha_4)} = 0$

of

b) $q_4 = 0$ (11.16)

of

c) $\left\{A_4+\int_0^{\alpha_4} tf(t)dt+\beta_4\int_{\alpha_4}^{\infty} f(t)dt-H(\alpha_1,\ldots,\alpha_4)+(\alpha_4-\beta_4)\left[1-F(\alpha_4)\right]\right\}= 0.$

De relatie (11.16a) is in strijd met (11.1). Immers

$$1-F(\alpha_4) = \int_{\alpha_4}^{\infty}\frac{1}{1000}\, e^{-\frac{t}{1000}}\, dt = e^{-\frac{\alpha_4}{1000}} \quad \text{en dus}$$

$$\frac{f(\alpha_4)}{1-F(\alpha_4)} = \frac{1}{1000} \neq 0 \quad \text{voor iedere waarde van } \alpha_4.$$

De relatie (11.16b) is gelijkwaardig met de uitspraak: "minstens één van de waarden α_i $(i=1,2,3)$ is onbegrensd". Een dergelijke keuze kan niet optimaal zijn.

Na deze overwegingen blijft dus over de relatie (11.16c). Uit (11.13c) en (11.16c) volgt onmiddellijk:

$$\alpha_i = \beta_i + \left\{H_i(\alpha_1,\ldots,\alpha_4) - H(\alpha_1,\ldots,\alpha_4)\right\}\frac{\sum_{h=1}^{i} q_h}{q_i\, F(\alpha_i)} \quad \text{voor } i \leq 3$$

$$\alpha_4 = \beta_4 + \frac{H(\alpha_1,\ldots,\alpha_4)-\left\{A_4+\int_0^{\alpha_4} tf(t)dt+\beta_4\int_{\alpha_4}^{\infty} f(t)dt\right\}}{1-F(\alpha_4)} \tag{11.17}$$

De relaties (11.17) bieden ons 4 vergelijkingen met 4 onbekenden, nl. $\alpha_1,\ldots,\alpha_4$. De optimale waarden van α_i zijn helaas niet op directe wijze te bepalen. Wij trachten daarom met behulp van een iteratie-procedure de optimale waarden van α_i te benaderen. Daartoe kiezen wij een willekeurig stel waarden $\alpha_i^{(0)}$ $(i=1,\ldots,4)$ en substitueren deze waarden voor α_i in de rechterleden van (11.17). De op deze wijze verkregen numerieke waarden van de rechterleden van (11.17) noemen wij $\alpha_i^{(1)}$. Vervolgens substitueren wij deze waarden in de rechterleden van (11.17) en verkrijgen op analoge wijze de waarden $\alpha_i^{(2)}$. Zo voortgaande construeren wij vier getallenrijen $\{\alpha_i^{(n)}\}$ $(i=1,\ldots,4)$. Indien ieder van deze getallenrijen convergeert tot een limiet α_i^*, dan voldoen de vier waarden α_i^* aan de relaties (11.17).

Deze iteratieprocedure behoeft niet altijd tot een oplossing te leiden. Kiezen wij echter $\alpha_i^{(0)} = A_{i+1} - A_1 + \beta_i$, dan blijkt dat de rijen $\{\alpha_i^{(n)}\}$ convergeren. De onderstaande tabel geeft de waarden van $\alpha_i^{(n)}$ voor n=0,1,2,3,4,5.

Tabel 11.I

Waarden van $\alpha_i^{(n)}$

n \ i	1	2	3	4
0	1050,00	1100,00	1100,00	1050,00
1	1309,10	1465,80	1475,82	1425,82
2	1319,30	1478,32	1489,87	1439,86
3	1319,39	1478,01	1489,32	1439,32
4	1319,49	1478,55	1489,87	1439,87
5	1319,16	1478,23	1489,54	1439,54

De berekeningen zijn steeds met zes decimalen uitgevoerd. Men kan nu bewijzen dat de functie $H(\alpha_1,\alpha_2,\alpha_3,\alpha_4)$ inderdaad minimaal is voor de volgende (afgeronde) waarden van α_i^*:

$$\left.\begin{aligned} \alpha_1^* &= 1319 \\ \alpha_2^* &= 1478 \\ \alpha_3^* &= 1490 \\ \alpha_4^* &= 1440. \end{aligned}\right\} \qquad (11.18)$$

Verder kan men laten zien dat uit (11.17) volgt

$$\alpha_4^* - \beta_4 = \alpha_3^* - \beta_3 , \qquad (11.19)$$

waarmee we het resultaat (11.18) kunnen controleren. Inderdaad geldt 1440 - 450 = 1490 - 500.

Uit (11.18) volgt tenslotte dat een schade

in de eerste klasse wordt geclaimd als zij groter is dan *f* 1.319.--,
in de tweede klasse wordt geclaimd als zij groter is dan *f* 1.478.--,
in de derde klasse wordt geclaimd als zij groter is dan *f* 1.490.--,
in de vierde klasse wordt geclaimd als zij groter is dan *f* 1.440.--.