STICHTING
MATHEMATISCH CENTRUM
2e BOERHAAVESTRAAT 49
AMSTERDAM

AFDELING MATHEMATISCHE STATISTIEK

Rapport S265 (C13)

Leergang Besliskunde

Hoofdstuk XIV

Eénstaps- en meerstaps-beslissingsproblemen

door

G. de Leve
en
J. van de Lune

3e druk
februari 1963

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 Pure Research (Z.W.O.) and the Central National Council for Applied Scientific Research in the Netherlands (T.N.O.), by the Municipality of Amsterdam and by several industries.

1. Inleiding

In dit hoofdstuk zullen wij onderscheid maken tussen één-staps- en meerstapsbeslissingsproblemen.

Men spreekt van een éénstapsbeslissingsprobleem indien slechts op één tijdstip een beslissing behoeft te worden genomen. Bij een meerstapsbeslissingsprobleem daarentegen wordt van de beslisser verwacht, dat hij beslissingen zal nemen op een groter aantal opeenvolgende tijdstippen (bijv. elke maandagmorgen om 9 uur).

Voorbeelden van éénstapsbeslissingsproblemen vindt men in hoofdstuk XII (Lineaire Programmering). Het is duidelijk, dat in voorbeeld 1.1 (van hoofdstuk XII) het beslissingsprobleem is opgelost, zodra de optimale legering is vastgesteld.

Voorbeelden van meerstapsbeslissingsproblemen vindt men in hoofdstuk IX, paragraaf 7. Immers, als de gegevens in tabel 1.I (hoofdstuk XII) van week tot week variëren, dan kunnen wij spreken van een meerstapsbeslissingsprobleem, omdat elke week opnieuw de optimale samenstelling van het mengsel bepaald dient te worden. Dit meerstapsbeslissingsprobleem kan omgekeerd ock gezien worden als een aantal <u>afzonderlijke</u> één-stapsbeslissingsproblemen.

Men kan nu de naam meerstapsbeslissingsprobleem reserveren voor al die problemen, waarvoor geldt dat de te nemen beslissingen <u>niet onafhankelijk</u> van elkaar kunnen worden bepaald. Maar ook deze beperking levert geen ondubbelzinnige indeling van de problemen. Immers het reeds als éénstapsbeslissingsprobleem aangemerkte voorbeeld (1.1) van hoofdstuk XII kan ook nog op een andere manier geformuleerd worden als een meerstapsbeslissingsprobleem. Wij behoeven slechts te stellen, dat de beslisser <u>achtereenvolgens</u> met tussenpozen van b.v. één minuut de grondstoffen A, B, D en E in de mengketels doet en dat hij op die tijdstippen beslist hoeveel van de betreffende grondstof hij zal gebruiken. Op de opéénvolgende tijdstippen worden dus beslissingen genomen, die in verband met de gestelde eisen niet onafhankelijk zijn.

Omgekeerd kunnen volledig deterministische meerstapsbeslissingsproblemen altijd herleid worden tot éénstapsbeslissingsproblemen. Immers bij een volledig deterministisch beslissingsprobleem weet men van te voren precies hoe het één en ander zich in de toekomst zal gaan ontwikkelen. Het is dus niet nodig om met het vaststellen van de te nemen beslissingen te wachten tot de desbetreffende beslissingstijdstippen zijn aangebroken Men kan op het eerste beslissingstijdstip dus al aangeven, welke reeks van beslissingen optimaal zal zijn. Dit betekent, dat men op één tijdstip één beslissing neemt en wel een beslissing, waarvan de componenten de beslissingen zijn die achtereenvolgens moeten worden uitgevoerd.

Men zou hieruit kunnen concluderen, dat men de eigenlijke meerstapsbeslissingsproblemen slechts vindt onder de stochastische beslissingsproblemen. Immers bij een stochastisch meerstapsbeslissingsprobleem kan men de toekomstige ontwikkelingen niet met zekerheid vaststellen. Aangezien de keuze van de beslissing veelal afhangt van de toestand waarin zij wordt genomen, moet men derhalve met het vaststellen van de optimale beslissing wachten tot het betreffende beslissingstijdstip is aangebroken.

Men kan nu echter tegenwerpen, dat de oplossing van het stochastische meerstapsbeslissingsprobleem wordt gegeven door een (optimale) strategie [1]. Aan de hand van deze strategie wordt op ieder beslissingstijdstip de optimale beslissing genomen. Het beslissingsprobleem is dus opgelost zodra de optimale strategie is vastgesteld. Dit dient te geschieden op het eerste beslissingstijdstip en derhalve zou men kunnen stellen dat op het eerste beslissingstijdstip een "super beslissing" (strategie) wordt bepaald, waaruit t.z.t. de te nemen beslissingen ondubbelzinnig volgen.

Dit impliceert, dat ook het stochastische meerstapsbeslissingsprobleem geformuleerd kan worden als een éénstapsbeslissingsprobleem.

Uit bovenstaande beschouwingen volgt, dat in principe alle meerstapsbeslissingsproblemen vertaald kunnen worden in éénstapsbeslissingsproblemen en dat ze daarom als zodanig kunnen worden opgelost.

----

1) In deze cursus is een strategie een voorschrift dat op elk beslissingstijdstip aan elke mogelijk optredende toestand een beslissing toevoegt.

Deze procedure wordt dan ook meer dan eens toegepast. Deze benaderingswijze is vooral dan succesvol, wanneer men de mogelijke strategiën kan vastleggen met behulp van een eindig aantal parameters .

Is dit het geval, dan wordt, wiskundig gezien, de oplossing van het meerstapsbeslissingsprobleem vaak op dezelfde wijze verkregen als die van een éénstapsbeslissingsprobleem.

Hoewel men vaak door de keuze van een andere formulering een éénstapsbeslissingsprobleem in een meerstapsbeslissingsprobleem kan overvoeren, blijkt in de praktijk dat het ene beslissingsprobleem zich eenvoudiger laat oplossen als éénstapsbeslissingsprobleem, terwijl het andere gemakkelijker oplosbaar is als het gesteld wordt als meerstapsbeslissingsprobleem. Wij zullen derhalve beide formuleringen afzonderlijk behandelen.

2. Het éénstapsbeslissingsprobleem

2. 1. Het niet-stochastische beslissingsprobleem

Elk lineair programmeringsprobleem is, zoals reeds in de inleiding werd vermeld, een voorbeeld van een éénstapsbeslissingsprobleem. Laat een functie z_0 gegeven zijn, welke lineair afhangt van de twee variabelen x_1 en x_2:

$$z_0=c_1x_1+c_2x_2$$

waarbij c_1 en c_2 bekende getallen zijn.

De vraag luidt: Voor welke waarden van x_1 en x_2, die niet negatief zijn en bovendien aan de ongelijkheden:

$$\begin{aligned} a_{11}x_1+a_{12}x_2 &\leqq b_1 \\ a_{21}x_1+a_{22}x_2 &\leqq b_2 \end{aligned} \qquad (b_i \geqq 0)$$

is de functie z_0 maximaal?

Indien wij de volgende notaties invoeren

$$S=(c_1,c_2)$$
$$X=(x_1,x_2) \qquad (2.1)$$
$$H(S;X)= S.X= c_1.x_1+c_2.x_2$$

en indien wij het gebied van de toegelaten beslissingen in de beslissingsruimte D aangegeven met D_o dan kan bovenstaand probleem ook als volgt geformuleerd worden:
Bepaal die X uit D_o, waarvoor de functie

$$z_o = H(S,X) \qquad (2.2.)$$

maximaal is.

In deze laatste formulering wordt niet geëist dat de functie (2.2) lineair is; vervolgens hebben wij geen uitspraak gedaan over de vorm van het gebied D_o van de toegelaten beslissingen. De laatste formulering is dan ook veel algemener en zal derhalve gebruikt worden om alle éénstapsbeslissingsproblemen te formuleren.

In 2.1.1 zullen wij een voorbeeld geven van een éénstapsbeslissingsprobleem.

2.1.1 Een inkoopprobleem

Een fabriek heeft voor een speciale opdracht een grote hoeveelheid entabogeen nodig. Entabogeen kan op elk tijdstip terplaatse worden ingekocht in iedere gewenste hoeveelheid à *f* 5000,-- per kg..

Men kan ook jaarlijks (op 1 juli) entabogeen kopen in Centraal Afrika en deze per boot laten verzenden naar de fabriek. Indien men een hoeveelheid van x kg. inkoopt, dan komen de totale kosten, inclusief vervoer, op

$$3000x + 20x^2 \qquad (2.3)$$

guldens [1].

1) "Grote hoeveelheden kunnen slechts ten koste van veel inspanning worden verkregen. Vandaar dat de inkoop-kosten per eenheid

$$\frac{3000x+20x^2}{x} = 3000+20x \qquad (2.4)$$

groter worden naarmate de gehele bestelde hoeveelheid x groter wordt."

Veronderstel vervolgens, dat

a) de fabriek het benodigde geld moet lenen à 8% per jaar.
b) het werk wordt aangevangen direct na aankomst van de grondstoffen uit Afrika.
c) het werk negen maanden zal duren.
d) het verbruik van entabogeen constant zal zijn en dat men in totaal 60 kg. nodig heeft.
e) de opdrachtgever zal betalen zodra het project wordt opgeleverd

Gevraagd wordt nu hoeveel kg. entabogeen in Afrika en hoeveel kg. ter plaatse moet worden ingekocht.

<u>Oplossing</u>

Stel de in Afrika in te kopen hoeveelheid entabogeen gelijk aan $x \leqq 60$kg.. De inkoopkosten bedragen dan

$$3000x+20x^2 \qquad (2.5)$$

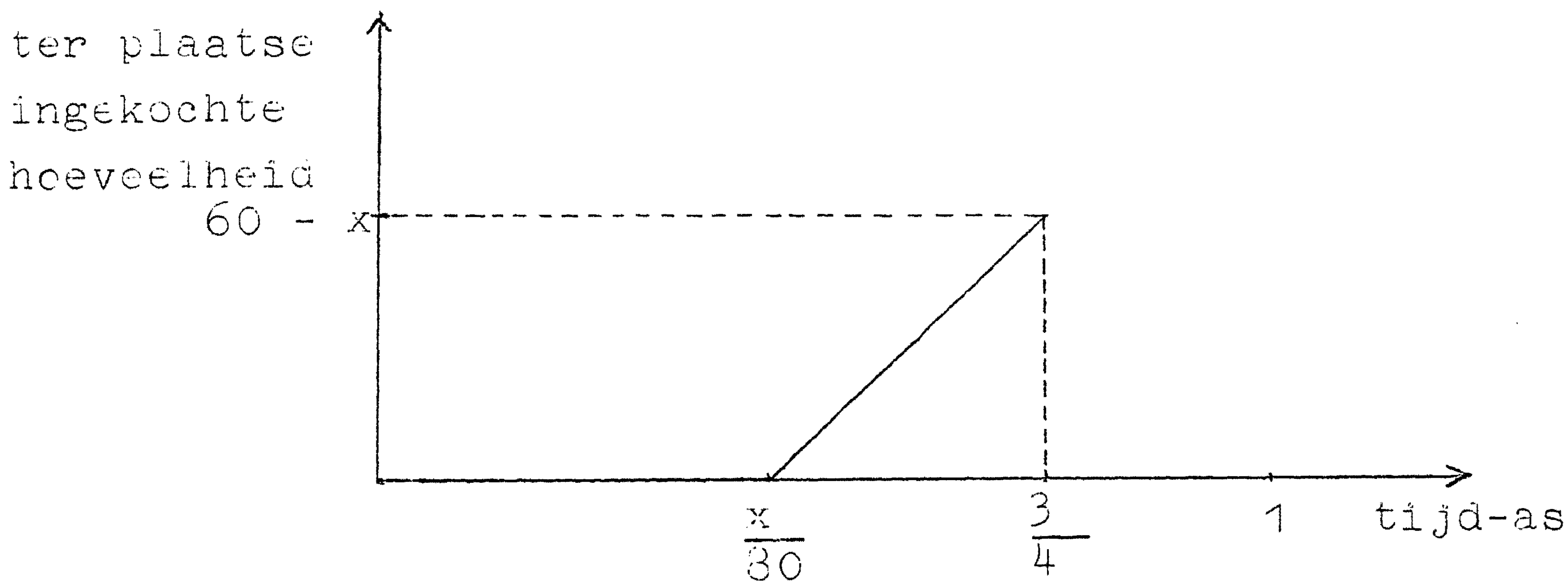
guldens.

Dit bedrag moet worden geleend van de bank. Als gevolg van rentederving zal een bedrag van

$$(1+\frac{9}{12}\ .0,08)\ [3000x+20x^2] \qquad (2.6)$$

aan de bank terugbetaald moeten worden. De verbruikssnelheid van entabogeen is 60 kg. in 9 maanden, d.w.z. $\frac{12}{9}$.60 kg.=80 kg. per jaar. De in Afrika ingekochte hoeveelheid is dus op na $\frac{x}{80}$ jaar. Gedurende de overige periode van $\frac{(60-x)}{80}$ jaar wordt voortdurend entabogeen in zeer kleine hoeveelheden ingekocht op de plaatselijke markt

Fig. 1 geeft van deze laatstgenoemde inkoop een overzicht.



fig, 1.

Overzicht van de op de plaatselijke markt ingekochte hoeveelheid entabogeen.

De inkoopkosten in de naaste omgeving bedragen

$$(60-x)5000 \text{ guldens} \qquad (2.7)$$

Indien de totale hoeveelheid (60-x) kg. wordt ingekocht op het tijdstip $\frac{x}{80}$ dan bedraagt de rente van het totale bedrag (2.7)

$$0{,}08 \cdot \frac{60-x}{80} \cdot (60-x) \cdot 5000 \text{ guldens} \qquad (2.8)$$

Door de "continue inkoop" wordt, zoals eenvoudig kan worden ingezien de rente gegeven door:

$$\frac{1}{2} \cdot 0.08 \cdot \frac{60-x}{80} \cdot (60-x) \cdot 5000 = 2{,}5\ (60-x)^2 \text{ guldens} \qquad (2.9)$$

Indien een hoeveelheid x wordt ingekocht in Afrika, dan worden de totale kosten gegeven door:

$$1.06\ \left[3000x+20x^2\right] + (60-x) \cdot 5000 + 2{,}5(60-x)^2 \qquad (2.10)$$

Na uitwerking is (2.10) gelijk aan:

$$\begin{aligned} &23{,}7x^2-2120x+309.000 = \\ &=23{,}7\ (x-44{,}7257)^2+261590{,}7987. \end{aligned} \qquad (2.11)$$

Deze functie is minimaal voor x = 44,7257 en het minimum bedraagt 261.590,7987.

In dit vraagstuk was het gebied D_o van de toegelaten beslissingen identiek met de beslissingsruimte D. Aan de keuze van de beslissing werd geen enkele bijzondere beperking opgelegd.

Veronderstel nu dat men in Afrika alleen partijen kan inkopen welke groter dan of gelijk aan 5o kg. zijn. Het gebied D_o bestaat nu niet uit de punten x die voldoen aan $x \geqq 50$ en uit het punt x=0.

Uit (2.11) volgt, dat, als voor de optimale x geldt: $x > 0$, men voor x moet kiezen x=50. De totale onkosten worden dan gegeven door:

$$23{,}7(50-44{,}7257)^2+261590{,}7987=262250{,}091. \qquad (2.12)$$

Voor x=0 vinden wij voor de totale kosten *f* 309.000. Bijgevolg wordt de optimale waarde van x gegeven door x=50.

2.2. Het stochastische éénstapsbeslissingsprobleem

Bij een stochastisch beslissingsprobleem kunnen wij het toekomstige verloop van het systeem in de toestandsruimte niet met zekerheid voorspellen. We kunnen in verband hiermee ook niet van te voren vaststellen wat het effect zal zijn van een te nemen beslissing. De kosten, welke in de betreffende tijdsperiode zullen worden gemaakt zijn stochastisch, d.w.z. zij bezitten een kansverdeling.

Bij een groot aantal beslissingsproblemen is het zinvol om de verwachting van de kosten in de beschouwing te betrekken. Men dient zich er echter steeds van bewust te zijn dat de werkelijke kosten achteraf anders kunnen zijn. Herhaalt het beslissingsprobleem zich in de loop van de tijd en kiest men steeds dezelfde beslissing, dan zullen bij een groot aantal herhalingen de gemiddelde onkosten met een grote mate van zekerheid dicht in de buurt liggen van de verwachting van de kosten. Deze verwachting is derhalve een zeer voor de hand liggend kriterium voor de optimale strategie.

3. Meerstapsbeslissingsproblemen.

Ter vereenvoudiging van de komende discussie zullen wij eerst een aatal begrippen omschrijven.

Het eerste begrip, dat wij invoeren heet systeem. Het systeem is het object van de beschouwingen. Aangenomen wordt dat het systeem zich in verschillende toestanden kan bevinden en dat deze toestanden beschreven kunnen worden m.b.v. een aantal kwantitatieve grootheden. Een toestand kan dus worden weergegeven door een rij van getallen.

Een dergelijke rij van getallen noemt men een toestands-vector S.

Verder stellen wij vast, dat als de toestanden van een systeem beschreven kunnen worden met behulp van m kwantitatieve grootheden, iedere toestand weergegeven kan worden door een punt in een m-dimensionaal Cartesiaans assenstelsel. Deze ruimte wordt toestandsruimte genoemd.

Bij de problemen welke wij in deze paragraaf zullen beschouwen moet op een aantal discrete tijdstippen een beslissing worden genomen.

Indien bij een probleem op n discrete tijdstippen een beslissing dient te worden genomen, dan spreekt men van een N-stapsbeslissingsprobleem. Zodra de eerste beslissing is genomen gaat het n-stapsbeslissingsprobleem over in een (N-1)-stapsbeslissingsprobleem.

Indien bij een probleem op oneindig veel discrete tijdstippen een beslissing genomen dient te worden, dan spreekt men van een ∞-stapsbeslissingsprobleem. Zodra de eerste beslissing is genomen gaat het ∞-stapsbeslissingsprobleem over in een nieuw ∞-stapsbeslissingsprobleem.

Verder stellen wij vast, dat ook een beslissing X kan worden vastgelegd met behulp van een aantal kwantitatieve grootheden.

Een dergelijke rij van getallen zullen wij een beslissingsvector noemen.

Voor het geval men de beslissingen kan aanduiden met behulp van r kwantitatieve grootheden correspondeert met iedere beslissing een punt X in de zgn. r-dimensionale beslissingsruimte D.

Bij een zorgvuldige beschouwing van meer stapsbeslissingsproblemen valt het ons op, dat de toestand van het systeem op het volgende beslissingstijdstip sterk kan worden beïnvloed door de toestand op het huidige beslissingstijdstip en door de op dit tijdstip te nemen beslissing.

Bij een volledig deterministisch beslissingsprobleem kan men met zekerheid vaststellen welke toestand het systeem op het volgende beslissingstijdstip zal aannemen als men tenminste de toestand en de beslissing op het huidige beslissingstijdstip kent.

Bij een beslissingsprobleem met stochastisch verloop kan men niet met zekerheid voorspellen wat de gevolgen zullen zijn van een beslissing in één of andere toestand. De toestand op het eerstvolgende beslissingstijdstip is dus onbekend, maar wel zal er een kansverdeling gedefinieërd zijn voor de betreffende toestand. Wij zullen hier steeds aannemen dat de gedaante van deze kansverdeling volledig bepaald zal zijn door de toestand en de beslissing op het huidige beslissingstijdstip. Het één en ander zal in paragraaf 3.2.1 met een voorbeeld worden toegelicht.

Verder kunnen wij constateren, dat er tengevolge van een beslissing steeds sprake is van een opbrengst (c.q. onkosten). Er zijn problemen, waarbij de opbrengst stochastisch is; in die gevallen beschouwt men de verwachting van de opbrengst.

De omvang van deze opbrengst (of de verwachting er van) zal o.m. afhangen van de genomen beslissing en de toestand waarin deze beslissing genomen werd. Wij zullen steeds aannemen dat de opbrengst alleen afhankelijk is van deze twee factoren.

Een reeks van beslissingen zullen wij optimaal noemen als de totale opbrengst maximaal is [1].

1) Er zijn problemen, waarbij niet de totale opbrengst maximaal gemaakt moet worden, maar een gewogen som van de opbrengsten.
Indien de opbrengst een verlies is dan wordt uiteraard geminimaliseerd.

Verder blijkt, dat wij voor de bepaling van de optimale beslissing op één tijdstip niet kunnen volstaan met alleen maar de opbrengst te beschouwen, welke een direct gevolg zal zijn van deze beslissing.

Immers deze beslissing bepaalt mede de toestand op het volgende beslissingstijdstip en dientengevolge de toekomstige opbrengsten.

Om het effect van een beslissing op toekomstige opbrengsten te kunnen nagaan, dient men, omdat de opbrengsten hierdoor mede worden bepaald, de toekomstige beslissingen te kennen.

3.1 Het niet-stochastische N stapsbeslissingsprobleem.

Bij een N stapsbeslissingbeslissingsprobleem begint men niet met de bepaling van de eerte optimale beslissing, maar met die van de laatste. Dat betekent, dat men een één-stapsbeslissingsprobleem moet oplossen.

Indien wij de directe opbrengst als gevolg van de j^{de} beslissing aangegeven met $h_j^{(N)}(S_j, X_j)$[1] dan zal men in de laatste stap trachten de functie $h_N^{(N)}(S_N, X_N)$ te maximaliseren naar X_N. Hiervoor dient men de toestand S_N van het systeem te kennen. Aangezien men deze toestand niet kent lost men het één-stapsbeslissingsprobleem op voor alle mogelijke toestanden. De maximaal te verkrijgen opbrengst

$$h_N^{(N)}(S_N) = \max_{X_N} \; h_N^{(N)}(S_N, X_N) \qquad (3.1)$$

is dus een functie van de toestand S_N op het laatste beslissingstijdstip.

In paragraaf 3 hebben wij vastgesteld, dat bij een volledig deterministisch beslissingsprobleem de toestand S_j op een beslissingstijdstip volledig wordt bepaald door de toestand S_{j-1} op het voorafgaande beslissingstijdstip en de daarop genomen beslissing X_{j-1}.

---

1) De bovenindex (N) geeft aan dat we te maken hebben met een N-staps beslissingsproblemen.

Bijgevolg kan de maximaal te verkrijgen opbrengst $h_N^{(N)}(S_N)$ in de laatste stap niet alleen geschreven worden als een functie van de toestand S_N op het laatste beslissingstijdstip maar ook als een functie $'h_{N-1}^{(N)}(S_{N-1},X_{N-1})$ van de toestand S_{N-1} en de beslissing X_{N-1} op het daaraan voorafgaande beslissingstijdstip.

Op het $(N-1)^{ste}$ beslissingstijdstip hebben wij dus twee opbrengsten en wel $'h_{N-1}^{(N)}(S_{N-1},X_{N-1})$, de opbrengst te verkrijgen in de laatste stap, en de directe opbrengst $h_{N-1}^{(N)}(S_{N-1},X_{N-1})$. Beide opbrengsten kunnen samengevoegd worden tot één opbrengst $^*h_{N-1}^{(N)}(S_{N-1},X_{N-1})$ waarvoor geldt:

$$^*h_{N-1}^{(N)}(S_{N-1},X_{N-1}) = h_{N-1}^{(N)}(S_{N-1},X_{N-1}) + {}'h_{N-1}^{(N)}(S_{N-1},X_{N-1}). \tag{3.2}$$

Na deze bewerkingen houden wij een (N-1)-stapsbeslissingsprobleem over met de opbrengsten:

$$h_j^{(N-1)}(S_j,X_j) = h_j^{(N)}(S_j,X_j) \quad \text{voor } j=1,\dots,N-2$$

en
$$h_j^{(N-1)}(S_j,X_j) = {}^*h_{N-1}^{(N)}(S_{N-1},X_{N-1}) \quad \text{voor } j=N-1. \tag{3.3}$$

Wij gaan dit (N-1)-stapsbeslissingsprobleem op dezelfde wijze oplossen als het oorspronkelijke N-stapsbeslissingsprobleem, met als gevolg dat het (N-1)-stapsbeslissingsprobleem wordt teruggebracht tot een (N-2)-stapsbeslissingsprobleem. Indien wij op deze wijze voortgaan wordt uiteindelijk het N-stapsbeslissingsprobleem gereduceerd tot een één-stapsbeslissingsprobleem.

3.1.1 Een produktieprobleem

Een chemische industrie heeft een kontrakt afgesloten met één van zijn cliënten.

In het kontrakt staat, dat

a. de fabriek zich verplicht tot levering van

D_1 kg. abaraat op 1 februari 1964
D_2 kg. abaraat op 1 maart 1964
D_3 kg. abaraat op 1 april 1964

b. de fabriek op de hierboven genoemde tijdstippen ook het duurdere benesol mag leveren, maar dan tegen de voor abaraat vastgestelde prijs.

Benesol is altijd in voorraad, maar abaraat zal speciaal voor de cliënt moeten worden gefabriceerd.

Het produktieniveau van abaraat kan alleen aan het begin van iedere maand worden veranderd. Elke verandering van het produktieniveau brengt kosten met zich mee.
Wordt het produktieniveau omgeschakeld van P' naar P", dan worden de bijbehorende kosten gegeven door

$$\alpha_1 . (P' - P'')^2 \qquad (3.4)$$

Waarbij α_1 een bekende constante is (b.v. $\alpha_1 = 2$). Op de afleveringstijdstippen kunnen zich nu de volgende situaties voordoen:

a. de vraag naar abaraat is groter dan de aanwezige hoeveelheid. Dit betekent, dat het verschil V tussen de aanwezige en de gevraagde hoeveelheid abaraat negatief is. Als gevolg hiervan moet er dus benesol bijgeleverd worden.

b. de vraag naar abaraat is kleiner dan de aanwezige hoeveelheid. Het verschil V tussen de aanwezige en de gevraagde hoeveelheid abaraat zal nu positief zijn. Er blijft dus abaraat over.

Als de vraag en de productie niet aan elkaar gelijk zijn, dan zullen er extra kosten gemaakt worden.

In situatie a. worden deze extra kosten veroorzaakt door de aflevering van het duurdere benesol, terwijl in situatie b. door de beperkte houdbaarheid van abaraat het restant als verloren moet worden beschouwd. Aangenomen wordt nu dat de kosten in beide situaties kunnen worden weergegeven door één kostenfunctie:

$$\alpha_2 V^2 \qquad (3.5)$$

waarbij α_2 een bekende constante is (b.v. $\alpha_2=3$).

Als V negatief is dan geeft $\alpha_2 V^2$ de extra kosten aan in situatie a) en als V positief is dan geeft $\alpha_2 V^2$ de kostprijs aan van het overschot aan abaraat.

Gevraagd een produktieprogramma op te stellen, waarvoor de extra kosten minimaal zijn.

Oplossing

Het systeem in dit beslissingsprobleem is de productie van abaraat.

De toestand S_i van het systeem wordt gegeven door:

a) het beslissingstijdstip i (voor i=1 is dit 1 januari 1964),

b) het produktieniveau P_{i-1} van de vorige maand.

Zoals wij in 3.1 hebben vastgesteld beginnen wij onze berekeningen met het beschouwen van het laatste beslissingstijdstip.

Op het laatste beslissingstijdstip wordt de toestand van het systeem gegeven door:

a) i=3

b) het produktieniveau P_2 in februari 1964

Gevraagd wordt het produktieniveau P_3 te bepalen voor maart. De kosten van een produktieomschakeling van P_2 naar P_3 wordt gegeven door:

$$\alpha_1 (P_3 - P_2)^2 \qquad (3.6)$$

De restantvoorraad op 1 april wordt gegeven door P_3-D_3 als $P_3 \geq D_3$; als $P_3 \leq D_3$ dan is er een tekort van D_3-P_3. Dit tekort is dan gelijk aan de hoeveelheid te leveren benesol.

In de laatste stap worden dus de volgende kosten gemaakt:

$$h_3^{(3)} (S_3, X_3) = \alpha_1 (P_3 - P_2)^2 + \alpha_2 (P_3 - D_3)^2. \qquad (3.7)$$

De produktie P_3 moet nu zodanig worden gekozen dat (3.7) minimaal is.

Door toepassing van een bekende stelling uit de differentiaalrekening (het nul stellen van de afgeleide van (3.7) naar P_3 vinden we

$$P_3 = \frac{\alpha_1 P_2 + \alpha_2 D_3}{\alpha_1 + \alpha_2} = \frac{2 P_2 + 3 D_3}{5} \qquad (3.8)$$

en de kosten in de laatste stap bedragen

$$h_3^{(3)}(S_3) = \min_{X_3} h_3^{(3)}(S_3, X_3) = \frac{\alpha_1 \alpha_2}{\alpha_1 + \alpha_2} (P_2 - D_3)^2 =$$

$$= \frac{6}{5} (P_2 - D_3)^2 \qquad (3.9)$$

Vestigen we nu onze aandacht op het tweede beslissingstijdstip; de kosten te maken in de maand februari worden gegeven door

$$h_2^{(3)}(S_2, X_2) = 2 (P_2 - P_1)^2 + 3 (P_2 - D_2)^2. \qquad (3.10)$$

De kosten in de maand maart worden weergegeven door (3.9). De totale kosten te maken in februari en maart worden dus gegeven door

$$h_2^{(2)}(S_2, X_2) \overset{*}{=} h_2^{(3)}(S_2, X_2) = 2 (P_2 - P_1)^2 + 3 (P_2 - D_2)^2 + \frac{6}{5} (P_2 - D_3)^2, \qquad (3.11)$$

en deze kosten kunnen we opvatten als de kosten die gemaakt worden in de laatste stap van een 2-stapsbeslissingsprobleem. De vorm (3.11) is minimaal voor

$$P_2 = \frac{2P_1 + 3D_2 + 1{,}2D_3}{6{,}2}$$

en de minimale waarde bedraagt

$$h_2^{(2)}(S_2) = \frac{4{,}2P_1^2 + 4{,}8\, D_2^2 + 3D_3^2 - 6P_1D_2 - 2{,}4P_1D_3 - 3{,}6D_2D_3}{3{,}1} \qquad (3.12)$$

Ten slotte berekenen we het optimale produktieniveau , P_1 voor de maand januari.
De kosten voor januari bedragen 1)

$$h_1^{(3)}(S_1,X_1) = h_1^{(2)}(S_1,X_1) =$$

$$= 2\,(P_1-P_o)^2 + 3(P_1-D_1)^2 = 2P_1^2+3(P_1-D_1)^2 \qquad (3.13)$$

Door het probleem verder op te vatten als een één-stapsbeslissings-probleem kunnen we de totale kosten als volgt schrijven

$$h_1^{(1)}(S_1,X_1) = {}^{*}h_1^{(2)}(S_1,X_1) = h_1^{(2)}(S_1,X_1) + {}'h_1^{(2)}(S_1,X_1) =$$

$$= 2P_1^2+3(P_1-D_1)^2 + \frac{4,2P_1^2+4,8D_2^2+3D_3^2-6P_1D_2-2,4P_1D_3-3,6D_2D_3}{3,1} =$$

$$= 6,355P_1^2-(6D_1+1,935D_2+0,774D_3)P_1+3D_1^2+1,548D_2^2+0,968D_3^2+$$

$$-\ 1,161D_2D_3 \qquad (3.14)$$

De vorm (3.14) is minimaal voor

$$P_1 = 0,472D_1+0,152D_2+0,061D_3. \qquad (3.15)$$

Wanneer we nu "terug gaan rekenen" dan vinden we voor de optimale productieniveaus

$$P_1^{*} = 0,472\ D_1+0,152\ D_2+0,061\ D_3$$

$$P_2^{*} = 0,152\ D_1+0,533\ D_2+0,213\ D_3$$

---

1) We stellen het productieniveau P_o van eind december 1963 hier gelijk aan nul.

$$P_3^* = 0{,}061\ D_1 + 0{,}?13\ D_2 + 0{,}685\ D_3$$

en de mimimale kosten bedragen

$$1{,}584\ D_1^2 + 1{,}401\ D_2^2 + 0{,}944\ D_3^2 - 0{,}914\ D_1D_2 - 0{,}365\ D_1D_3 - 1{,}279\ D_2D_3.$$

Het bovenstaande produktieprobleem hebben wij zojuist opgelost als zijnde een meerstapsbeslissingsprobleem; maar omdat het probleem deterministisch van aard is kunnen we het ook opvatten als een éénstapsbeslissingsprobleem en als zodanig oplossen (zie paragraaf 1).

2e oplossing

We stellen de produktieniveaus in de maanden januari, februari en maart respectievelijk x,y en z en bepalen de optimale beslissingsvector (x^*, y^*, z^*).
De totale kosten bedragen [1].

$$\alpha_1(x-P_0)^2+\alpha_2(x-D_1)^2+\alpha_1(x-y)^2+\alpha_2(y-D_2)^2+\alpha_1(y-z)^2+\alpha_2(z-D_3)^2 \tag{3.16}$$

als we de beslissing (x,y,z) nemen.
In (3.16) staat een kwadratische functie f (x,y,z) van de drie variabelen x,y en z; met behulp van de differentiaalrekening kan worden aangetoond dat deze functie slechts één uiterste waarde heeft en wel een minimum.
We vinden dit minimum door de partiële afgeleiden

$$f'_x\ (x,y,z),\quad f'_y\ (x,y,z)\ \text{en}\ f'_z\ (x,y,z)$$

alle gelijk aan nul te stellen.
Dit leidt tot de volgende betrekkingen

$$f'_x(x,y,z)=2\alpha_1(x-P_0)+2\alpha_2(x-D_1)+2\alpha_1(x-y)=0$$

1) Het produktieniveau P_0 van december 1963 stellen wij nu niet gelijk aan nul.

$$f'_y\,(x,y,z) = -2\alpha_1(x-y) + 2\alpha_2(y-D_2) + 2\alpha_1\,(y-z) = 0 \qquad (3.17)$$

$$f'_z\,(x,y,z) = -2\alpha_1(y-z) + 2\alpha_2(z-D_3) = 0$$

Dit stelsel vergelijkingen laat zich op zeer eenvoudige wijze omvormen tot

$$\begin{aligned} (2\alpha_1+\alpha_2)x - \alpha_1\,y \qquad\qquad &= \alpha_2\,D_1 + \alpha_1\,P_o \\ -\alpha_1 x + (2\alpha_1+\alpha_2)y - \alpha_1\,z &= \alpha_2\,D_2 \qquad (3.18) \\ -\alpha_1\,y + (\alpha_1+\alpha_2)\,z &= \alpha_2\,D_3\,. \end{aligned}$$

Nemen we, evenals in het voorgaande, aan dat α_1 en α_2 respectievelijk de waarden 2 en 3 hebben en dat $P_o=0$, dan gaat (3.18) over in

$$\begin{aligned} 7\,x - 2\,y \qquad\quad &= 3\,D_1 \\ -2\,x + 7\,y - 2\,z &= 3\,D_2 \qquad (3.19) \\ -2\,y + 5\,z &= 3\,D_3 \end{aligned}$$

en de oplossing hiervan luidt

$$x^* = \frac{93\,D_1 + 30\,D_2 + 12\,D_3}{197}$$

$$y^* = \frac{30\,D_1 + 105\,D_2 + 42\,D_3}{197} \qquad (3.20)$$

$$z^* = \frac{12\,D_1 + 42\,D_2 + 135\,D_3}{197}$$

waarmee we de optimale waarden van P_1, P_2 en P_3 gevonden hebben. Substitueren we deze waarden in (3.16) dan vinden we voor de minimale kosten

$$\frac{61464}{38809} D_1^2 + \frac{54372}{38809} D_2^2 + \frac{36645}{38809} D_3^2 +$$

$$- \frac{35460}{38809} D_1D_2 - \frac{14184}{38809} D_1D_3 - \frac{49644}{38809} D_2D_3 \quad .$$

Deze oplossing is dezelfde als de vorige, zoals de lezer desgewenst zelf kan nagaan.
De totale produktie bedraagt in dit geval

$$x^* + y^* + z^* = \frac{135\ D_1 + 177\ D_2 + 189\ D_3}{197} < D_1+D_2+D_3 \ ,$$

waaruit volgt dat de totale produktie kleiner is dan de totale vraag.
Tellen we vergelijkingen in (3.18) op dan vinden we

$$(\alpha_1+\alpha_2)x + \alpha_2 y + \alpha_2 z = \alpha_2(D_1+D_2+D_3) + \alpha_1\ P_o \ . \qquad (3.21)$$

De oplossing (x^*, y^*, z^*) van het stelsel (3.18) moet ook voldoen aan (3.21); substitutie van deze oplossing in (3.21) levert

$$x^* + y^* + z^* = D_1 + D_3 + \frac{\alpha_1}{\alpha_2}\ (P_o - x^*) \qquad (3.22)$$

waaruit volgt dat in het algemene geval de totale produktie groter is dan de totale vraag als $P_o > x^*$; de totale produktie is kleiner dan de totale vraag als $P_o < x^*$ en vraag en produktie zijn aan elkaar gelijk als $P_o = x^*$.

3.1.2. Een opleidingsprobleem.

Veronderstel dat een koloniaal gebied wordt voorbereid op zelfbestuur, hetwelk over vijf jaar zal worden verleend. De beherende mogendheid wil in de nog resterende jaren trachten een gezonde basis te scheppen voor de onafhankelijkheid. Daartoe zijn verschillende plannen uitgewerkt, waarbij één betrekking heeft op de gezondheidszorg. In het door deskundigen samengestelde rapport kan men lezen, dat het wenselijk is drie cursussen te organiseren en wel die opleiden tot

a) ziekenverzorger
b) instructeur-ziekenverzorger A
c) instructeur-ziekenverzorger B.

Alle cursussen duren één jaar en kunnen slechts in de hierboven gegeven volgorde worden doorlopen. De uit te zenden blanke instructeurs kunnen alle opleidingen geven, terwijl een instructeur-ziekenverzorger A alleen kan opleiden tot ziekenverzorger en een instructeur ziekenverzorger B de bevoegdheid bezit leerlingen op te leiden zowel tot ziekenverzorger als tot instructeur-ziekenverzorger A.

In de opleiding tot ziekenverzorger kan iedere instructeur maximaal vijf leerlingen instrueren.

In de opleiding tot instructeur ziekenverzorger A kan iedere instructeur maximaal drie leerlingen opleiden.

In de opleiding tot instructeur ziekenverzorger B kan elke blanke instructeur maximaal twee leerlingen opleiden.

Aangenomen wordt dat het leerlingenreservoir onbeperkt is en dat iedere leerling geschikt is om elke opleiding te volgen; verder wordt voorgeschreven dat, wanneer iemand eenmaal de status van instructeur bereikt heeft, hij nooit meer als ziekenverzorger dienst zal doen. De deskundigen vragen zich in het rapport af, hoe de instructeurs in de verschillende jaren over de opleidingen moeten worden verdeeld om het maximale aantal "verplegingsjaren" te verkrijgen in de nog resterende vijf jaren van het beheer, als verder gegeven is dat er voor dit doel vijf blanke instructeurs beschikbaar zijn. Een "verplegingsjaar" is een jaar, waarin één ziekenverzorger gedurende één jaar geheel ter beschikking van de verpleging staat. Men spreekt b.v. van 3 verplegingsjaren als

4 ziekenverzorgers gedurende 2 jaar geheel ter beschikking van de verpleging staan.

Omschrijving van het mathematisch model.

Onder het systeem zullen wij verstaan "de opleiding van verplegingspersoneel". Alle personen die bij dit systeem betrokken zijn kunnen worden ingedeeld naar hun status σ_i en naar hun functie φ_j.

Overzicht van de diverse status.

Status-aanduiding	Omschrijving van de status
σ_0	nog nooit bij een opleiding betrokken geweest
σ_1	alleen opgeleid tot ziekenverzorger
σ_2	opgeleid tot instructeur-ziekenverzorger A, maar niet tot instructeur -ziekenverzorger B
σ_3	opgeleid tot instructeur-ziekenverzorger B
σ_4	blanke instructeur

(3.23)

Overzicht van de mogelijke functies.

Functie-aanduiding	Omschrijving van de werkzaamheden
φ_0	in opleiding
φ_1	ziekenverzorger
φ_2	instructeur bij de opleiding van σ_0 tot σ_1
φ_3	instructeur bij de opleiding van σ_1 tot σ_2
φ_4	instructeur bij de opleiding van σ_2 tot σ_3

(3.24)

Schema voor het verdelen der functies over de verschillende status.

Status	Mogelijke functie.
σ_0	φ_0
σ_1	φ_0, φ_1
σ_2	φ_0, φ_2
σ_3	φ_2, φ_3
σ_4	$\varphi_2, \varphi_3, \varphi_4$

(3.25)

Wanneer we zeggen dat iemand de functie φ_j bekleedt, dan wil dat zeggen dat deze persoon daadwerkelijk te werk gesteld is in de functie φ_j, ongeacht de status van de betreffende persoon; het is b.v. heel goed mogelijk dat iemand, ook al heeft hij de status σ_4, gedurende één of meer jaren te werk gesteld is in de functie φ_2. Met behulp van (3.23) definiëren wij de toestandsvectoren S_1, S_2, S_3, S_4 en S_5. De toestandsvector S_k ($k=1,2,\ldots,5$) zal aangeven hoeveel personen er op het k- de beslissingstijdstip (nog voordat de k- de beslissing genomen is) bij het systeem betrokken zijn en in welke status.

Wij schrijven deze toestandsvectoren als volgt

$$\left.\begin{aligned} S_1 &= \{s_{0,1}, s_{1,1}, s_{2,1}, s_{3,1}, s_{4,1}\} \\ S_2 &= \{s_{0,2}, s_{1,2}, s_{2,2}, s_{3,2}, s_{4,2}\} \\ &- - - - - - - - - - - - - - - - - - \\ S_5 &= \{s_{0,5}, s_{1,5}, s_{2,5}, s_{3,5}, s_{4,5}\} \end{aligned}\right\} \quad (3.26)$$

zodat de toestandsvector S_k wordt weergegeven door

$$S_k = \{s_{0,k}, s_{1,k}, s_{2,k}, s_{3,k}, s_{4,k}\} \; ; \qquad (3.26a)$$

hierbij geeft de component $s_{i,k}$ het aantal personen aan dat op het k- de beslissingstijdstip de status σ_i heeft. Omdat $s_{o,k}$ voor $k=1,2,\ldots, 5$ steeds oneindig groot is, kunnen we deze component gemakshalve uit S_k weglaten. Hierdoor gaat (3.26a) over in

$$S_k = \left\{s_{1,k},\ s_{2,k},\ s_{3,k},\ s_{4,k}\right\} \qquad (3.27)$$

Wanneer we dit formalisme strikt handhaven, dan zullen we moeten stellen dat

$$S_1 = \left\{0,\ 0,\ 0,\ 5\right\} \qquad (3.28)$$

omdat, zolang het systeem nog geen jaar in werking is, er nog niemand is met de status σ_i ($i=1,\ 2,3$), terwijl gegeven is dat er 5 blanke instructeurs beschikbaar zijn. Het ligt voor de hand om nu reeds vast te stellen dat er in het eerste jaar zoveel mogelijk inheemsen moeten worden opgeleid tot ziekenverzorger.
Het gevolg hiervan is dat

$$S_2 = \left\{25,\ 0,\ 0,\ 5\right\}. \qquad (3.29)$$

Het verdere verloop van het systeem is afhankelijk van de beslissingen die op het tweede, derde, vierde en vijfde beslissingstijdstip genomen zullen worden.
De te nemen beslissingen definiëren we als vectoren met componenten x_{ijk} ($i = 0,1,2,3,4;\ \ j = 0,1,2,3,4;\ \ k = 1,2,3,4,5.$), in een of andere volgorde gerangschikt; hierbij is x_{ijk} het aantal personen dat op het k-de beslissingstijdstip de status σ_i heeft en in het k-de jaar te werk gesteld zal worden in de functie φ_j.

Eerste oplossing.

We zullen het probleem eerst oplossen als zijnde een 5-stapsbeslissingsprobleem.
De opbrengst (het aantal verplegingsjaren) in het 5-de jaar bedraagt

$$h_5^{(5)}\ (S_5,\ X_5) = s_{1,5} - x_{10\,5}, \qquad (3.30)$$

en dus

$$h_5^{(5)}\ (S_5) = \max_{X_5}\ h_5'^{(5)}\ (S_5,\ X_5) =$$

$$= \max_{X_5} (s_{1,5} - x_{105}) = s_{1,5}. \qquad (3.31)$$

Bovenstaand resultaat is vrijwel evident omdat het voor de totale opbrengst alleen maar nadelig is om personen met de status σ_1 in het vijfde jaar nog verder op te leiden; we kiezen dus

$$x_{105} = 0 . \qquad (3.32)$$

We hebben het 5-stapsbeslissingsprobleem nu herleid tot een 4-staps beslissingsprobleem.

$$h_4^{(4)}(S_4, X_4) = {}^{*}h_4^{(5)}(S_4, X_4) =$$

$$= h_4^{(5)}(S_4, X_4) + {}'h_4^{(5)}(S_4, X_4) =$$

$$= (s_{1,4} - x_{104}) + s_{1,5} = (s_{1,4} - x_{104}) + (s_{1,4} + x_{004} - x_{104}) =$$

$$= 2s_{1,4} + x_{004} - x_{104} . \qquad (3.32)$$

Daar het opleiden van personen met de status σ_1 in het vierde jaar bij voorbaat ongunstig is, kiezen we

$$x_{104} = 0 . \qquad (3.33)$$

Zo heeft het evenmin zin om in het vierde jaar nog op te leiden van σ_2 tot σ_3. Met behulp hiervan leiden we gemakkelijk af.

$$h_4^{(4)}(S_4) = \max_{X_4} h_4^{(4)}(S_4, X_4) =$$

$$= \max_{X_4} (2\, s_{1,4} + x_{004}) = 2\, s_{1,4} + \max_{X_4} x_{004} =$$

$$= 2\, s_{1,4} + 5\, (s_{2,4} + s_{3,4} + s_{4,4}) . \qquad (3.34)$$

Uit het voorgaande volgt dat er in het vierde jaar alleen aandacht besteed zal worden aan de opleiding van σ_0 tot σ_1.
Op het derde beslissingstijdstip zijn er nog geen personen met de status σ_3 zoals blijkt uit het volgende schema

Beslissings-tijdstip	Mogelijk voorkomende status
1	σ_0, σ_4
2	σ_0, σ_1, σ_4
3	σ_0, σ_1, σ_2, σ_4

(3.35)

Leiden we in het derde jaar op van σ_2 tot σ_3 dan zullen er op het vierde beslissingstijdstip personen zijn met de status σ_3. In het vierde jaar zullen alleen instructeurs betrokken zijn bij de opleiding van σ_0 tot σ_1, en om instructeur bij deze opleiding te kunnen zijn is het voldoende de status σ_2 te hebben. Hieruit volgt dat het geen zin heeft om in het derde jaar (blanke) instructeurs beschikbaar te stellen voor de opleiding van σ_2 tot σ_3. Hieruit volgt dat er op het vierde beslissingstijdstip ook geen personen met de status σ_3 zullen zijn;

dus $$s_{3,4} = 0 \text{ (en } x_{203} = 0\text{)}. \qquad (3.36)$$

Substitutie van (3.36) in (3.34) levert

$$h_4^{(4)}\ (S_4) = 2\ s_{1,4} + 5\ (s_{2,4} + s_{4,4}) \qquad (3.37)$$

en omdat $s_{4,k} = 5$ voor $k = 1,2,3,4$ en 5 vinden we tenslotte

$$h_4^{(4)}\ (S_4) = 25 + 2\ s_{1,4} + 5\ s_{2,4}\ .$$

Hiervoor kunnen we schrijven

$$h_4^{(4)}(S_4) = {}'h_3^{(4)}(S_3,X_3) = 25 + 2(s_{1,3} + x_{003} - x_{103}) + 5(s_{2,3} + x_{103} + -x_{203}) =$$

$$= 25 + 2\ s_{1,3} + 5\ s_{2,3} + 2x_{003} + 3\ x_{103} \qquad (3.38)$$

want $$x_{203} = 0. \qquad (3.36)$$

Verder vinden we

$$h_3^{(3)}(S_3^1, X_3) = {}^{*}h_3^{(4)}(S_3, X_3) = h_3^{(4)}(S_3,X_3) + {}'h_3^{(4)}(S_3,X_3) =$$

$$= (s_{1,3} - x_{103}) + (25 + 2\ s_{1,3} + 5\ s_{2,3} + 2\ x_{003} + 3\ x_{103}) =$$

$$= 25 + 3\ s_{1,3} + 5\ s_{2,3} + 2\ x_{003} + 2x_{103} \qquad (3.39)$$

en dus $h_3^{(3)}(S_3) = \max\limits_{X_3} h_3^{(3)}(S_3, X_3) =$ (3.40)

$$= 25 + 3\ s_{1,3} + 5\ s_{2,3} + 2 \max_{X_3} (x_{003} + x_{103}).$$

Als er in het derde jaar x_{103} personen met de status σ_1 worden opgeleid tot σ_2 dan zijn hiervoor $\left[\frac{x_{103}}{3}\right]^+$ instructeurs nodig[1]; aangezien er in het derde jaar niemand wordt opgeleid van σ_2 tot σ_3 zullen er voor de opleiding van σ_0 tot σ_1

$$5 + s_{2,3} - \left[\frac{x_{103}}{3}\right]^+ \qquad (3.41)$$

instructeurs beschikbaar zijn. Het zal duidelijk zijn dat elk der laatstgenoemde instructeurs 5 leerlingen zal moeten opleiden.

Dus $$x_{003} = 5 \left\{5 + s_{2,3} - \left[\frac{x_{103}}{3}\right]^+\right\} \qquad (3.42)$$

Formule (3.40) kan dus ook geschreven worden als

$$h_3^{(3)}(S_3) = 25 + 3\ s_{1,3} + 5\ s_{2,3} + 2 \max_{X_3} \left\{25 + 5\ s_{2,3} + \right.$$
$$\left. - 5 . \left[\frac{x_{103}}{3}\right]^+ + x_{103}\right\} =$$

$$= 75 + 3\ s_{1,3} + 15\ s_{2,3} + 2 \max_{X_3} \left\{x_{103} - 5. \left[\frac{x_{103}}{3}\right]^+\right\}. \qquad (3.43)$$

Voor positieve waarden van x_{103} is de vorm tussen accoladen in (3.43) altijd negatief, terwijl deze vorm gelijk aan nul is als we

$$x_{103} = 0 \qquad (3.44)$$

kiezen.
Voor (3.43) kunnen we dus schrijven

$$h_3^{(3)}(S_3) = 75 + 3\ s_{1,3} + 15\ s_{2,3}. \qquad (3.45)$$

1) $\left[\frac{x_{103}}{3}\right]^+$ is het kleinste gehele getal dat groter dan of gelijk aan $\frac{x_{103}}{3}$ is.

Voorts is

$$h_2^{(2)}(S_2,X_2) = {}^{*}h_2^{(3)}(S_2,X_2) = h_2^{(3)}(S_2,X_2) + {}'h_2^{(3)}(S_2,X_2) =$$

$$= (s_{1,2} - x_{102}) + \{75 + 3\,(s_{1,2} + x_{002} - x_{102}) +$$

$$+ 15\,(s_{2,2} + x_{102} - x_{202})\}$$

maar daar $s_{2,2} = 0$ en $x_{202} = 0$, kunnen we hiervoor ook schrijven

$$h_2^{(2)}(S_2, X_2) = 75 + 4\,s_{1,2} + 3\,x_{002} + 11\,x_{102}. \tag{3.46}$$

Dus:

$$h_2^{(2)}(S_2) = \max_{X_2} \{75 + 4\,s_{1,2} + 3\,x_{002} + 11\,x_{102}\} =$$

$$= 75 + 4\,s_{1,2} + \max_{X_2}(3\,x_{002} + 11\,x_{102}). \tag{3.47}$$

Als er in het tweede jaar x_{102} personen worden opgeleid van σ_1 tot σ_2 dan zijn hiervoor

$\left[\frac{x_{102}}{3}\right]^{+}$ (blanke) instructeurs nodig;

hieruit volgt vrijwel direct dat dan x_{002} gelijk aan

$$x_{002} = 5\left\{5 - \left[\frac{x_{102}}{3}\right]^{+}\right\} \tag{3.48}$$

gekozen zal moeten worden.
We vinden zo

$$h_2^{(2)}(S_2) = 75 + 4\,s_{1,2} + \max_{X_2}\left\{75 - 15\left[\frac{x_{102}}{3}\right]^{+} + 11\,x_{102}\right\} =$$

$$= 150 + 4.s_{1,2} + \max_{X_2}\left\{11\,x_{102} - 15\left[\frac{x_{102}}{3}\right]^{+}\right\}. \tag{3.49}$$

Nu mag x_{102} nooit groter zijn dan de maximale opleidingscapaciteit van de vijf beschikbare blanke instructeurs, terwijl x_{102} tevens kleiner dan of gelijk aan het aantal personen met de status σ_1 op het tweede beslissingstijdstip moet zijn; x_{102} moet dus voldoen aan de voorwaarden

$$\left.\begin{array}{l} 0 \leqq x_{102} \leqq 5.\,3 = 15 \\ 0 \leqq x_{102} \leqq s_{1,2} \end{array}\right\} . \qquad (3.50)$$

Aangezien $s_{1,2} = 25$, zijn de twee ongelijkheden in (3.50) gelijkwaardig met de éne ongelijkheid

$$0 \leqq x_{102} \leqq 15.$$

Door de vorm die in (3.49) tussen accoladen staat uit te rekenen voor $x_{102} = 1,2,\ldots,$ 15 vinden we, dat

$$x_{102} = 15 \qquad (3.51)$$

de optimale keuze is voor x_{102}, met als gevolg

$$h_2^{(2)}\ (S_2) = 150 + 4\ s_{1,2} + 90 = 240 + 4\ s_{1,2}\,. \qquad (3.52)$$

Daar het onmiddellijk duidelijk is dat $s_{1,2} = 25$, is de maximale opbrengst dus

$$H^{*} = 240 + 4.25 = 340\ . \qquad (3.53)$$

Tenslotte geven we hieronder het optimale beslissingsschema op overzichtelijke wijze weer.

Beslissings-tijdstip	Optimale beslissing
1	Elke blanke instructeur moet 5 inheemsen opleiden van σ_0 tot σ_1.
2	Elke blanke instructeur moet 3 inheemsen opleiden van σ_1 tot σ_2 terwijl er 10 inheemsen als ziekenverzorger te werk gesteld moeten worden.
3	De opleiding van σ_1 tot σ_2 moet worden stopgezet terwijl elke beschikbare instructeur (in totaal zijn dat er intussen 20) 5 inheemsen moet opleiden van σ_0 tot σ_1. Er blijven 10 inheemsen als ziekenverzorger werkzaam.
4	Elke beschikbare instructeur moet 5 inheemsen opleiden van σ_0 tot σ_1. De overige 110

	inheemsen, die de status σ_1 hebben, zullen alle een taak als ziekenverzorger toegewezen krijgen.
5.	Alle personen met status σ_1 (210 man) moeten als ziekenverzorger te werk worden gesteld. Het verdere opleidingsschema is irrelevant.

Betreffende de aantallen verplegingsjaren geven wij nog het volgende overzicht

Jaar	Aantal verplegingsjaren
1	0
2	10
3	10
4	100 + 10 = 110
5	100+100+10 = 210
totaal	340

3.2. Het stochastische N-stapsbeslissingsprobleem

Ook in het stochastische N-stapsbeslissingsprobleem begint men niet met de bepaling van de eerste optimale beslissing, maar met die van de laatste. Dit houdt in dat eerst een stochastisch éénstapsbeslissingsprobleem wordt opgelost.

Indien wij de <u>verwachting</u> van de directe opbrengst van de j-de beslissing aangegeven met $h_j^{(N)}(S_j;X_j)$ dan zal men in de laatste stap trachten de functie $h_N^{(N)}(S_N,X_N)$ te maximaliseren naar X_N. Hiervoor dient men evenwel de toestand S_N van het systeem te kennen. Aangezien men deze toestand niet kent lost men het éénstapsbeslissingsprobleem op voor alle mogelijke toestanden.
De maximaal te verkrijgen opbrengst

$$h_N^{(N)}(S_N) = \max_{X_N} h_N^{(N)}(S_N,X_N) \qquad (3.65)$$

is dus een functie van de toestand S_N op het laatste beslissingstijdstip.
In paragraaf 3 hebben wij vastgesteld, dat bij een stochastisch beslissingsprobleem de kansverdeling van de toestand S_j op het j^{de} beslissingstijdstip volledig wordt bepaald door de toestand S_{j-1} en de beslissing X_{j-1}.

De opbrengst $h_N^{(N)}(S_N)$ is dus, gezien vanuit het $(N-1)^{ste}$ beslissingstijdstip, stochastisch. Indien wij dus op het $(N-1)^{ste}$ beslissingstijdstip rekening willen houden met de laatste beslissing, dan zullen wij de verwachting van de optimale opbrengst in de laatste stap in onze beschouwingen moeten betrekken.

Deze verwachting, $\mathcal{E} h_N^{(N)}(\underline{S}_N)$ [1] kan nu wél worden opgevat als een functie van de toestand S_{N-1} en de beslissing X_{N-1}; deze functie zullen wij aangeven met $'h_N^{(N)}(S_{N-1},X_{N-1})$.

Als we de opbrengsten van de beide laatste stappen samenvoegen en deze som aangeven met $^{*}h_{N-1}^{(N)}(S_{N-1},X_{N-1})$ dan geldt:

$$^{*}h_{N-1}^{(N)}(S_{N-1},X_{N-1}) = h_{N-1}^{(N)}(S_{N-1},X_{N-1}) + {'h}_{N-1}^{(N)}(S_{N-1},X_{N-1}) \qquad (3\ 66)$$

---

1) Deze verwachting dient berekend te worden onder de voorwaarde dat op het (N-1)-ste beslissingstijdstip het systeem in de toestand S_{N-1} verkeerde en X_{N-1} de (N-1)-ste beslissing was.

Na deze bewerkingen houden wij een (N-1)-stapsbeslissingsprobleem over met de opbrengstfunctie

$$h_j^{(N-1)}\ (S_j,X_j) = h_j^{(N)}\ (S_j,X_j) \quad \text{voor } j=1,2,\ldots,N-2 \quad \text{en}$$

$$h_j^{(N-1)}\ (S_j,X_j) = {}^{*}h_{N-1}^{(N)}\ (S_{N-1},X_{N-1}) \quad \text{voor } j=N-1.$$

Zo doorgaande wordt het oorspronkelijke N-stapsbeslissingsprobleem herleid tot een éénstapsbeslissingsprobleem (vergelijk paragraaf (3.1).

3.2.1 Een verkoopprobleem

Een reisbureau heeft voor een periode van 6 jaar een hotel gepacht in een wintersportcentrum. Met een kolenhandelaar daar ter plaatse is een kontrakt afgesloten, waarin wordt bepaald dat de kolenhandelaar elk jaar een vaste hoeveelheid brandstof zal leveren tegen betaling van een vast bedrag B per jaar.

Verder is overeengekomen dat het reisbureau, in geval van ontevredenheid over de leveranties, aan het eind van elk jaar het kontrakt éénzijdig mag opzeggen.

De kolenhandelaar verkoopt drie soorten kolen:

(1) superkolen
(2) kwaliteitskolen
(3) huishoudkolen.

Levert de kolenhandelaar gedurende een jaar de kolensoort (i) dan bedraagt zijn winst a_i (i=1,2,3).

De kans op opzegging van het kontrakt na levering van de kolensoort (i) duiden we aan met p_i en deze kans zal onafhankelijk zijn van het jaar van levering.

De kolenhandelaar, een harde zakenman, vraagt zich nu af welke kolensoorten hij in de 6 opeenvolgende jaren bij het hotel zal afleveren.

Omschrijving van het mathematisch model

In dit geval zullen wij onder het systeem verstaan de kolen leverantie (of het kontrakt). Als het kontrakt is opgezegd dan zullen we dit uitdrukken door te zeggen dat het systeem zich in de toestand 0 bevindt; als het kontrakt niet is opgezegd dan

zullen we zeggen dat het systeem zich in de toestand 1 bevindt. De toestandsruimte bestaat hier dus uit slechts twee punten. De kolenhandelaar moet achtereenvolgens 6 maal een beslissing nemen en daarmee staat hij voor een 6-stapsbeslissingsprobleem. De beslissingsruimte D zal per definitie bestaan uit de vier punten 0,1,2 en 3. De beslissing X=0 komt overeen met het leveren van geen enkele kolensoort; deze beslissing wordt om vanzelfsprekende redenen dan en slechts dan genomen als het kontrakt is opgezegd (d.w.z. als S=0).
De beslissing X=i (i=1,2,3) houdt in dat gedurende het betreffende jaar de kolensoort (i) geleverd zal worden. De laatst genoemde beslissingen kunnen dus alleen genomen worden als S=1.
Het bovenstaande kan worden weergegeven in het volgende schema

Toestand S	Toegelaten beslissingen X
0	0
1	1,2,3

Oplossing
Als de kolenhandelaar in het zesde jaar mag leveren (dus als $S_6=1$ dan zal de zesde beslissing X_6^* zodanig gekozen moeten worden dat

$$h_6^{(6)}(1,X_6) \qquad (3.67)$$

maximaal is voor $X_6=y^*_6$.
Onderstaande tabel geeft een overzicht van de waarden die $h_6^{(6)}(1,X_6)$ kan aannemen voor alle mogelijke waarden van X_6.

X_6	$h_6^{(6)}(1,X_6)$
1	a_1
2	a_2
3	a_3

(3.68)

Voor de optimale beslissing X_6^* geldt dus:

$$h_6^{(6)}(1,X_6^*) = h_6^{(6)}(1) = \max_i \left[a_i\right] , \qquad (3,69)$$

wat wil zeggen dat de kolenhandelaar in het laatste jaar (als hij mag leveren) die kolensoort zal leveren waarop hij de grootste winst maakt.

Als de kolenhandelaar in het zesde jaar niet mag leveren (dus als $S_6=0$) dan zal hij zonder enige twijfel de beslissing $X_6=0$ nemen. Dezelfde redenering geldt voor alle beslissingstijdstippen waarop geldt $S=0$. Voor zo'n tijdstip geldt dan dat de verwachting van de nog te maken winst gelijk is aan nul, want als het kontrakt eenmaal is opgezegd worden er geen kolen meer geleverd voor het resterende deel van de 6 jaar.

Als de kolenhandelaar in het vijfde jaar mag leveren dan zal hij de vijfde beslissing X_5^* zodanig moeten kiezen dat

$$h_5^{(5)}(1,X_5) = {}^*h_5^{(6)}(1,X_5) = h_5^{(6)}(1,X_5) + {}'h_5^{(6)}(1,X_5) \qquad (3,70)$$

maximaal is voor $X_5=X_5^*$.

In onderstaande tabel geven we de waarden aan van de in (3,70) voorkomende functies, voor alle mogelijke waarden van X_5.

X_5	$h_5^{(6)}(1,X_5)$	${}'h_5^{(6)}(1,X_5)$
1	a_1	$(1-p_1) . h_6^{(6)}(1)$
2	a_2	$(1-p_2) . h_6^{(6)}(1)$
3	a_3	$(1-p_3) . h_6^{(6)}(1)$

Voor X_5^* moet dus gelden:

$$h_5^{(5)}(1,X_5^*) = h_5^{(5)}(1) = \max_i \left[a_i+(1-p_i) . h_6^{(6)}(1)\right] \qquad (3,71)$$

en uit deze betrekking kan X_5^* worden opgelost zodra de numerieke waarden van a_i en p_i $(i=1,2,3)$ bekend zijn.
Als de kolenhandelaar in het vierde jaar mag leveren dan moet de vierde beslissing X_4^* zodanig gekozen worden dat

$$h_4^{(4)}(1,X_4)={}^*h_4^{(5)}(1,X_4)=h_4^{(5)}(1,X_4)+\ 'h_4^{(5)}(1,X_4) \qquad (3,72)$$

maximaal is voor $X_4=X_4^*$.

Onderstaande tabel geeft een overzicht van de waarden die de functies in (3.72) kunnen aannemen.

X_4	$h_4^{(5)}(1,X_4)$	$'h_4^{(5)}(1,X_4)$
1	a_1	$(1-p_1)\ h_5^{(5)}(1)$
2	a_2	$(1-p_2)\ h_5^{(5)}(1)$
3	a_3	$(1-p_3)\ h_5^{(5)}(1)$

De beslissing X_4^* moet dus voldoen aan:

$$h_4^{(4)}(1,X_4^*)=h_4^{(4)}(1) = \max_i \left[a_i+(1-p_i).\ h_5^{(5)}(1)\right] \qquad (3.73)$$

en uit deze betrekking kunnen we X_4^* oplossen
Zo doorgaande vinden we dat de optimale beslissingen X_j^* $(j=1,2,3,4,5)$ moeten voldoen aan

$$h_j^{(j)}(1)=\max_i \left[a_i+(1-p_i)\ h_{j+1}^{(j+1)}(1)\right]. \qquad (3.74)$$

Van bovenstaand resultaat zullen wij nu een toepassing beschouwen.
Stel dat

$$\begin{array}{llll} a_1= & 435 & p_1= & 0,2 \\ a_2= & 790 & p_2= & 0,4 \\ a_3= & 1050 & p_3= & 0,6 \end{array} \qquad (3.75)$$

Uit (3.75) en (3.69) volgt $X_6^*=3$ en $h_6^{(6)}(1,X_6^*) = h_6^{(6)}(1) = 1050$.

Toepassing van (3.71) levert $X_5^*=3$, en $h_5^{(5)}(1,X_5^*) = 1470$, want

$$a_1+(1-p_1)\ h_6^{(6)}(1)=435 + 0{,}8 .\ 1050 = 1275$$

$$a_2+(1-p_2)\ h_6^{(6)}(1)=790 + 0{,}6 .\ 1050 = 1420 \qquad (3.76)$$

$$a_3+(1-p_3)\ h_6^{(6)}(1)=1050 + 0{,}4 .\ 1050 = 1470 .$$

Toepassing van (3.73) levert $X_4^*=2$ en $h_4^{(4)}(1,X_4^*)= h_4^{(4)}(1) = 1672$,

want $$a_1+(1-p_1)\ h_5^{(5)}(1)=435 + 0{,}8 .\ 1470 = 1611$$

$$a_2+(1-p_2)\ h_5^{(5)}(1)=790 + 0{,}6 .\ 1470 = 1672 \qquad (3.77)$$

$$a_3+(1-p_3)\ h_5^{(5)}(1)=1050+ 0{,}4 .\ 1470 = 1638 .$$

Verder vinden we: $X_3^*=2$ en $h_3^{(3)}(1) = 1793{,}20$, want

$$a_1+(1-p_1)\ h_4^{(4)}(1) = 435 + 0{,}8 .\ 1672 = 1772{,}60$$

$$a_2+(1-p_2)\ h_4^{(4)}(1) = 790 + 0{,}6 .\ 1672 = 1793{,}20 \qquad (3.78)$$

$$a_3+(1-p_3)\ h_4^{(4)}(1) = 1050 + 0{,}4 .\ 1672 = 1718{,}80$$

en $X_2^* = 1$ en $h_2^{(2)}\ (1,X_2^*) = h_2^{(2)}(1)= 1869{,}56$, want

$$a_1+(1-p_1)\ h_3^{(3)}(1) = 435 + 0{,}8 .\ 1793{,}20 = 1869{,}56$$

$$a_2+(1-p_2)\ h_3^{(3)}(1) = 790 + 0{,}6 .\ 1793{,}20 = 1865{,}92 \qquad (3.79)$$

$$a_3+(1-p_3)\ h^{(3)}(1) = 1050 + 0{,}4 .\ 1793{,}20 = 1767{,}38 .$$

Tenslotte vinden we $X_1^*=1$ en $h_1^{(1)}(1,X_1^*) = h_1^{(1)}(1) = 1930{,}65$, want

$$a_1+(1-p_1)\ h_2^{(2)}(1) = 435 + 0{,}8 .\ 1869{,}56 = 1930{,}65$$

$$a_2+(1-p_2)\ h_2^{(2)}(1) = 790 + 0{,}6 .\ 1869{,}56 = 1911{,}74 \qquad (3.80)$$

$$a_3+(1-p_3)\ h_2^{(2)}(1) = 1050 + 0{,}4 .\ 1869{,}56 = 1797{,}82 .$$

Samenvattende hebben we de oplossing weergegeven in onderstaande tabel.

Beslissingstijdstip j	Optimale beslissing X_j^*	Kolensoort die wordt afgeleverd in het j-de jaar
1	1	superkolen
2	1	superkolen
3	2	kwaliteitskolen
4	2	kwaliteitskolen
5	3	huishoudkolen
6	3	huishoudkolen

De optimale winstverwachting bedraagt 1930,65.

4. Het ∞-staps beslissingsprobleem.

In de vorige paragrafen hebben wij ons beperkt tot beslissingsproblemen, waarbij slechts een eindig aantal beslissingen genomen behoefde te worden.

In deze paragraaf zullen wij ons bezighouden met situaties waarin het aantal te nemen beslissingen onbegrensd is; het zal zonder meer duidelijk zijn dat de oplossing van deze problemen niet verkregen kan worden door (zoals we in de vorige paragraaf hebben gedaan) eerst de "laatste" beslissing optimaal te kiezen.

4.1 Het definiete ∞-staps beslissingsprobleem.

In dit geval blijkt uit de probleemstelling dat de toestand op het eerstvolgende beslissingstijdstip volledig bepaald wordt door de toestand op het huidige moment van beslissen en de op dit tijdstip genomen beslissing.
Verkeert het systeem op het j-de beslissingstijdstip in de toestand S_j en is de bijbehorende beslissing X_j, dan kan de toestand van het systeem op het (j+1)-ste beslissingstijdstip dus worden opgevat als een éénduidige functie

$$S_{j+1}= S_{j+1} (S_j,X_j)$$

van S_j en X_j.

In deze paragraaf zullen wij aannemen dat de relatie tussen S_{j+1} enerzijds en S_j en X_j anderzijds voor iedere waarde van j dezelfde is, m.a.w.

$$S_{j+1}=S(S_j,X_j) \ , \qquad (4.1)$$

Bovendien zullen we veronderstellen dat de opbrengstfunctie $h_j\ (S_j,X_j)$ voor alle beslissingstijdstippen dezelfde is; in plaats van $h_j\ (S_j,X_j)$ schrijven we in het vervolg dan ook $h(S_j,X_j)$. Onafhankelijk hiervan kunnen zich verder o.a. de volgende situaties voordoen

1. uit de probleemstelling volgt, dat bij de i-de beslissing de volledige opbrengst, te verkrijgen uit de toekomstige beslissingen, in rekening moet worden gebracht.

2. uit de probleemstelling volgt, dat op het i-de beslissingstijdstip slechts rekening wordt gehouden met een fractie

van de opbrengsten behorende bij de toekomstige beslissingen en wel zo dat de directe bijdrage van de j-de beslissing ($j \geqq 1$) tot de totale opbrengst gelijk is aan

$$\alpha^{j-1} h \quad (S_j, X_j) \tag{4.2}$$

Hierbij is $0 \leqq \alpha < 1$.

Bij de oplossing van een ∞-staps beslissingsprobleem nemen we eerst de eerste n beslissingstijdstippen in ogenschouw zonder ons te bekommeren om de vraag wat er op de volgende tijdstippen nog allemaal kan gebeuren. Op deze wijze hebben we een n-staps beslissingsprobleem geschapen waarvan we met de methoden van de vorige paragraaf de oplossing kunnen bepalen. De maximaal te verkrijgen opbrengst zal een functie $f_n(S_1)$ van de begintoestand S_1 zijn. Verder is gemakkelijk in te zien dat in geval 1, de betrekking

$$\begin{aligned} f_n(S_1) &= \max_{X_1} \left[h\ (S_1, X_1) + f_{n-1}\ (S_2) \right] = \\ &= \max_{X_1} \left[h\ (S_1, X_1) + f_{n-1}(S(S_1, X_1)) \right] \end{aligned} \tag{4.3}$$

zal gelden,

waarbij $f_{n-1}(S_2)$ gelijk is aan de maximale opbrengst vanaf het tweede beslissingstijdstip; omdat de toestand S_2 volkomen bepaald wordt door S_1 en X_1 kan $f_{n-1}(S_2)$ ook geschreven worden als een functie

$$f_{n-1}(S(S_1, X_1))$$

van een functie $S(S_1, X_1)$ van S_1 en X_1.

Voor geval 2 luiden deze formules

$$\begin{aligned} f_n(S_1) &= \max_{X_1} \left[h\ (S_1, X_1) + f_{n-1}\ (S_2) \right] = \\ &= \max_{X_1} \left[h\ (S_1, X_1) + f_{n-1}\ (S(S_1, X_1)) \right] \end{aligned} \tag{4.4}$$

Beschouwen wij nu de rij $f_n(S_1)$ dan kan het gebeuren dat deze rij voor elke vaste toestandsvector S_1 naar een getal $f(S_1)$ convergeert als n onbegrensd toeneemt. Bij elke toestandsvector S zullen we dan een limiet f(S) vinden en de verzameling limieten die zo ontstaat kunnen we opvatten als een functie op de verzameling van alle mogelijke begintoestanden. Als deze convergentie inderdaad optreedt dan mogen we verwachten dat de relaties (4.3) en (4.4) respectievelijk zullen overgaan in

$$f(S_1) = \max_{X_1} \left[h \ (S_1, X_1) + f(S_2) \right] =$$
$$= \max_{X_1} \left[h \ (S_1, X_1) + f(S(S_1, X_1)) \right] \qquad (4.5)$$

en

$$f(S_1) = \max_{X_1} \left[h \ (S_1, X_1) + \alpha f(S_2) \right] =$$
$$= \max_{X_1} \left[h \ (S_1, X_1) + \alpha f(S(S_1, X_1)) \right] \qquad (4.6)$$

De oplossing van het definiete ∞ -staps beslissingsprobleem kan nu bepaald worden zodra de functie $f(S)$ bekend is.
Want als $f(S)$ bekend is voor elke waarde van S dan kunnen we door middel van de betrekkingen (4.5) en (4.6) voor elke begintoestand S_1 de bijbehorende optimale beslissing X_1^* bepalen.

4.2 Het stochastische ∞ -staps beslissingsprobleem.[1)]

Nu blijkt uit de probleemstelling dat de kansverdeling van de toestandsvector behorende bij het eerstvolgende beslissingstijdstip volledig bepaald wordt door de toestand op het huidige moment van beslissen en de hierop genomen beslissing.
Verkeert het systeem op het j-de beslissingstijdstip in de toestand S_j en is de j-de beslissing X_j dan is de toestand op het (j+1)-ste beslissingstijdstip dus een stochastische grootheid $\underline{S}_{j+1}$ waarvan de kansverdeling volkomen bepaald wordt door S_j en X_j. Bovendien zullen wij aannemen dat deze kansverdeling niet afhankelijk is van het beslissingstijdstip. Als het systeem dus op b.v. het 5^e en het 9^e beslissingstijdstip dezelfde toestand S aanneemt en we nemen op beide beslissingstijdstippen dezelfde beslissing X, dan zal de kansverdeling van $\underline{S}_6$ dezelfde zijn als die van $\underline{S}_{10}$.
Ook bij het stochastische ∞ -staps beslissingsprobleem onderscheiden we twee gevallen

1. uit de probleemstelling volgt, dat bij de i-de beslissing de verwachting van de volledige opbrengst, te verkrijgen uit de toekomstige beslissingen, in rekening moet worden gebracht.

----

1) Paragraaf 4.1 kan beschouwd worden als een bijzonder geval van paragraaf 4.2

2. uit de probleemstelling volgt, dat op het i-de beslissingstijdstip slechts rekening wordt gehouden met een fractie van de opbrengsten[1] behorende bij de toekomstige beslissingen en wel zo dat de directe bijdrage van j-de beslissing ($j \geqq i$) tot de totale opbrengst gelijk is aan

$$\alpha^{j-i}\, h\,(S_j, X_j), \qquad (4.7)$$

met $0 \leqq \alpha < 1$.

Evenals in paragraaf 4.1 beschouwen wij nu eerst weer de eerste n beslissingstijdstippen.
De verwachting van de maximaal te verkrijgen opbrengst in deze n stappen zal ook hier een functie

$$f_n(S_1)$$

zijn van de toestand S_1 op het eerste beslissingstijdstip.
In geval 1 zal voor deze functie de betrekking

$$f_n(S_1) = \max \left[h\,(S_1, X_1) + \mathcal{E} f_{n-1}(\underline{S}_2) \right]^{2)} \qquad (4.8)$$

gelden, en in geval 2

$$f_n(S_1) = \max \left[h\,(S_1, X_1) + \alpha . \mathcal{E} f_{n-1}(\underline{S}_2) \right] \qquad (4.9)$$

Ook hier kan zich het geval voordoen dat de rij $f_n(S_1)$ voor elke toestandsvector S_1 convergeert naar een limiet $f(S_1)$. De limietfunctie $f(S)$ die we op deze wijze verkrijgen zal dan voldoen aan

----

1) Als ook deze directe opbrengsten stochastisch zijn dan nemen we hiervoor hun verwachtingen.

2) De verwachting $\mathcal{E} f_{n-1}(\underline{S}_2)$ dient te worden berekend onder de voorwaarde dat het systeem op het eerste beslissingstijdstip de toestand S_1 aannam en dat daarop de beslissing X_1 genomen werd. Deze verwachting wordt hierom dan ook vaak genoteerd als een voorwaardelijke verwachting

$$\mathcal{E} \left\{ f_{n-1}(\underline{S}_2) \;\middle|\; S_1, X_1 \right\} .$$

In deze laatste notatie komt duidelijker uit dat deze verwachting een functie is van S_1 en X_1.

$$f(S_1) = \max_{X_1} \left[h(S_1, X_1) + \mathcal{E} f(\underline{S}_2) \right] \qquad (4.10)$$

of aan

$$f(S_1) = \max_{X_1} \left[h(S_1, X_1) + \alpha \mathcal{E} f(\underline{S}_2) \right] \qquad (4.11)$$

(vergelijk de formules (4.8) en (4.9)) voor elke begintoestand S_1. De oplossing van het stochastische ∞-staps beslissingsprobleem kan nu bepaald worden zodra de functie f(S) bekend is. De optimale eerste beslissing X_1^* kan dan n.l. bepaald worden met behulp van (4.10) c.q. (4.11).

4.3. Voorbeeld van een ∞-staps beslissingsprobleem.

In dit voorbeeld willen wij nog eens het verkoopprobleem uit paragraaf 3.2.1. beschouwen, met dit verschil, dat het kontrakt niet voor 6 jaar zal gelden maar voor onbeperkte tijd; hiermee hebben wij het 6- stapsbeslissingsprobleem omgevormd tot een ∞ staps beslissingsprobleem.
Het proces van de kolenleverantie begint in de toestand 1; de bijbehorende optimale winstverwachting stellen we f(1).
Zou het proces in de toestand 0 zijn begonnen dan zou de (eveneens optimale) winstverwachting f(o) geweest zijn. Het gestelde probleem is nu op bijzonder eenvoudige wijze op te lossen met behulp van formule (4.10).
Het zal zonder meer duidelijk zijn dat f(o) gelijk is aan nul zodat we alleen nog maar de waarde van f(1) behoeven te bepalen met de optimale beslissing in de toestand 1.
Volgens (4.10) moet f(1) voldoen aan

$$\begin{aligned} f(1) &= \max_i \left\{ a_i + p_i \,.\, f(o) + (1-p_i) \,.\, f(1) \right\} = \\ &= \max_i \left\{ a_i + (1-p_i) \,.\, f(1) \right\}. \end{aligned} \qquad (4.12)$$

Uit (4.12) volgt gemakkelijk:

$$0 = \max_i \left\{ a_i - p_i \, f(1) \right\}. \qquad (4.13)$$

Als i^* de optimale beslissing is in de toestand 1 dan moet i^* voldoen aan de betrekking

$$a_{i^*} - p_{i^*} \,.\, f(1) = 0, \qquad (4.14)$$

waaruit voor f(1) volgt

$$f(1) = \frac{a_{i^*}}{p_{i^*}} . \tag{4.15}$$

Met het oog op (4.15) is gemakkelijk in te zien dat f(1) moet voldoen aan

$$f(1) = \max_i \frac{a_i}{p_i} . \tag{4.16}$$

Nu is

$$\frac{a_1}{p_1} = \frac{435}{0,2} = 2175$$

$$\frac{a_2}{p_2} = \frac{790}{0,4} = 1975$$

$$\frac{a_3}{p_3} = \frac{1050}{0,6} = 1750$$

waaruit volgt dat $i^* = 1$ en $f(1) = 2175$.

4.4. Oplossingsmethoden.

4.4.1. Eerste iteratie-methode.

Indien de rij van opbrengstfuncties $f_n(S)$ van de bijbehorende n-staps beslissingsproblemen voor iedere S naar de opbrengstfunctie f(S) convergeert, dan kan deze functie uiteraard iteratief worden bepaald.
Heeft men deze functie f(S) eenmaal gevonden dan kan men met behulp van één der relaties [(4.5), (4.6), (4.10), (4.11)] voor iedere begintoestand S_1 de optimale beslissing X_1^* bepalen.
Deze verzameling van optimale beslissingen kunnen we opvatten als een functie $X^*(S)$ van de begintoestand S. Dit beslissingsvoorschrift wordt de <u>optimale strategie</u> genoemd.

4.4.2. Tweede iteratiemethode.

Indien voor iedere <u>strategie</u> χ op eenvoudige wijze de bijbehorende opbrengstfunctie f(S) bepaald kan worden, dan kan men dikwijls ook op een andere iteratieve wijze de functionaalvergelijkingen [(4.5), (4.6), (4.10), (4.11)] oplossen.
Men begint met een "willekeurige" strategie $\chi^{(0)}(S)$ te kiezen en bepaalt daarbij de opbrengstfunctie $f^{(0)}(S)$. Uit de functie $f^{(0)}(S)$ leiden we nu een nieuwe strategie $\chi^{(1)}(S)$ af, op de wijze die in paragraaf 4.4.1 is aangegeven.

Bij de strategie $\chi^{(1)}(S)$ bepalen we weer de opbrengstfunctie $f^{(1)}(S)$ en hieruit leiden we weer de strategie $\chi^{(2)}(S)$ af enz. Zo ontstaat een rij van strategieën $\chi^{(i)}(S)$ en een rij opbrengstfuncties $f^{(i)}(S)$.
Indien deze rijen convergeren en men weet dat de oplossing van de corresponderende functionaalvergelijking [(4.5),(4.6),(4.10),(4.11)] bestaat en ondubbelzinnig is, dan kan men ook op deze wijze de optimale strategie $\chi^{*}(S)$ bepalen.
Indien de rij van opbrengstfuncties $f^{(i)}(S)$ van de bijbehorende n-stapsbeslissingsproblemen niet convergeert, dan kan men dus, ondanks de divergentie van de rij $f^{(i)}(S)$, toch de optimale strategie $\chi^{*}(S)$ bepalen.
Doorgaans wenst men slechts de optimale strategie te kennen en dan is hiermee het probleem opgelost.
Ook komt het wel voor dat de limiet

$$\lim_{i\to\infty} \frac{f^{(i)}(S)}{i} = \lim_{i\to\infty} g_i(S)=g(S) \qquad (4.17)$$

bestaat.
De functies $g_i(S)$ stellen de gemiddelde opbrengst per beslissing voor bij een i-staps beslissingsprobleem. Bij menig probleem convergeert deze rij van functies tot een functie $g(S)$.
Deze functie neemt in de eindoplossing dan de plaats in van de (niet bestaande) functie $f(S)$.

In par. 4.5 zullen wij d.m.v. een voorbeeld een methode bespreken, waarmee men de functionaalvergelijkingen op een meer directe wijze kan oplossen. Deze methode is slechts te gebruiken als de functionaalvergelijkingen van een bepaalde structuur zijn.

4.4.3 Een uitgebreide groep van ∞-staps beslissingsproblemen kan worden opgelost met behulp van de theorie der Markov-ketens. Bij deze problemen geldt dat het verloop van het systeem, bij toepassing van een vaste strategie χ, beschreven kan worden met behulp van een Markov-keten. Van deze beslissingsproblemen hebben wij reeds een drietal opgelost in hoofdstuk IX.
In paragraaf 4.5 zullen wij nog een andere oplossingsmethode, de "policy improvement" methode, bespreken.

4.5. Een ∞-staps beslissingsprobleem dat opgelost kan worden, zonder gebruik te maken van iteratie-procedures.

Een hartstochtelijke sportvisser heeft in een café horen spreken van een boer, die twee vennetjes bezit met bijzonder mooie en grote vissen; helaas heeft de boer het vissen in beide vennetjes verboden. In het ene vennetje (V_1) zouden zich x en in het andere (V_2) y vissen bevinden.
Alle sprekers in het café waren het er over eens dat de kans om bij vennetje V_1 betrapt te worden gelijk was aan p_1, terwijl deze kans voor V_2 gelijk zou zijn aan p_2. De sportvisser, die door deze mededelingen nieuwsgierig is geworden, besluit eens op verkenning uit te gaan. Op grond van zijn ruime ervaring meent hij na een vluchtige studie van de situatie te mogen vaststellen dat de fractie van het totaal aantal in V_1 resp. V_2 aanwezige vissen, die die hij per keer aan de haak kan slaan steeds gelijk zal zijn aan r_1 resp. r_2. Thuis gekomen overweegt hij, of hij als vreemdeling in deze omgeving geacht zal worden het visverbod te kennen.
Een ∞-staps beslissingsprobleem ontstaat, zodra de sportvisser besluit elke vrije zaterdag in één van de beide vennetjes te gaan vissen (totdat hij betrapt wordt), zich daarbij elke zaterdagmorgen opnieuw afvragend in welke ven hij zijn hengel zal uitwerpen. Het spreekt haast vanzelf dat het deze sportvisser er om te doen is zijn totale vangst zo groot mogelijk te maken.

Oplossing.

Als de visser slechts éénmaal gaat vissen, dan wordt de verwachte vangst gegeven door

$$f_1(x,y) = \max\left\{(1-p_1)r_1x\ ;\ (1-p_2)r_2y\right\} \qquad (4.18)$$

Uit deze relatie volgt onmiddellijk:

1°) $f_1(\alpha x,\alpha y) = \alpha f_1(x,y) \qquad (\alpha \geqq 0)$ (4.19)

2°) $f_1(x,y)$ is continu in x en y.

Indien hij tweemaal gaat vissen dan wordt de verwachting van de totale opbrengst gegeven door:

$$f_2(x,y) = \max\Big\{(1-p_1)r_1x + (1-p_1)f_1\left[(1-r_1)x,y\right]\ ; \qquad (4.20)$$

$$(1-p_2)r_2y + (1-p_2)f_1\left[x;(1-r_2)y\right]\Big\}.$$

Uit (4.20) volgt dat ook voor $f_2(x,y)$ geldt:

$$1^\circ)\quad f_2(\alpha x, \alpha y) = \alpha f_2(x,y) \qquad (\alpha \geqq 0) \tag{4.21}$$
$$2^\circ)\quad f_2(x,y) \text{ is continu in } x \text{ en } y.$$

Indien hij n maal gaat vissen dan wordt de verwachting van de totale vangst gegeven door:

$$f_n(x,y) = \max\Big\{(1-p_1)r_1x+(1-p_1)f_{n-1}\big[(1-r_1)x,y\big]\ ;\ (1-p_2)r_2y+(1-p_2)f_{n-1}\big[x;(1-r_2)y\big]\Big\}. \tag{4.22}$$

Met behulp van de methode der volledige inductie kan men voor elk geheel getal $n>0$ bewijzen:

$$1^\circ)\quad f_n(\alpha x,\alpha y) = \alpha f_n(x,y) \quad (\alpha \geqq 0) \tag{4.23}$$
$$2^\circ)\quad f_n(x,y) \text{ is continu in } x \text{ en } y.$$

Men kan bewijzen dat de limiet $f_n(x,y)$ voor $n \to \infty$ voor iedere waarde van x en y bestaat en dat de relatie (4.22) overgaat in [1])

$$f(x,y) = \max\Big\{(1-p_1)r_1x + (1-p_1)f\big[(1-r_1)x,y\big]\ ;\ (1-p_2)r_2y + (1-p_2)f\big[x;(1-r_2)y\big]\Big\}. \tag{4.24}$$

Deze functionaalvergelijking heeft één oplossing.

Ook voor de functie $f(x,y)$ geldt:

$$1^\circ)\quad f(\alpha x,\alpha y) = \alpha f(x,y) \quad (\alpha \geqq 0) \tag{4.25}$$
$$2^\circ)\quad f(x,y) \text{ is continu in } x \text{ en } y.$$

We geven de beslissing om in V_1 te gaan vissen aan met D_1. De alternatieve beslissing wordt met D_2 aangeduid. De keuze tussen beide beslissingen wordt bepaald door de toestand van het systeem op het beslissingstijdstip. Voor ieder punt (x,y) zou men na kunnen gaan welke beslissing optimaal is. Uit (4.25) volgt dat deze keuze ook optimaal is voor alle punten op de verbindingslijn met de oorsprong. Uit (4.25) volgt bovendien, dat als in een punt D_1 te verkiezen is boven D_2 dit dan ook geldt

----

1) Voor het bewijs van deze uitspraken wordt verwezen naar R. Bellman, Dynamic Programming blz. 64 e.v.

voor een klein genoeg gekozen omgeving van het punt. Wanneer men deze feiten combineert dan moet de toestandsruimte te verdelen zijn in sectoren, waarvan de vorm aangegeven is in fig. 4.1.

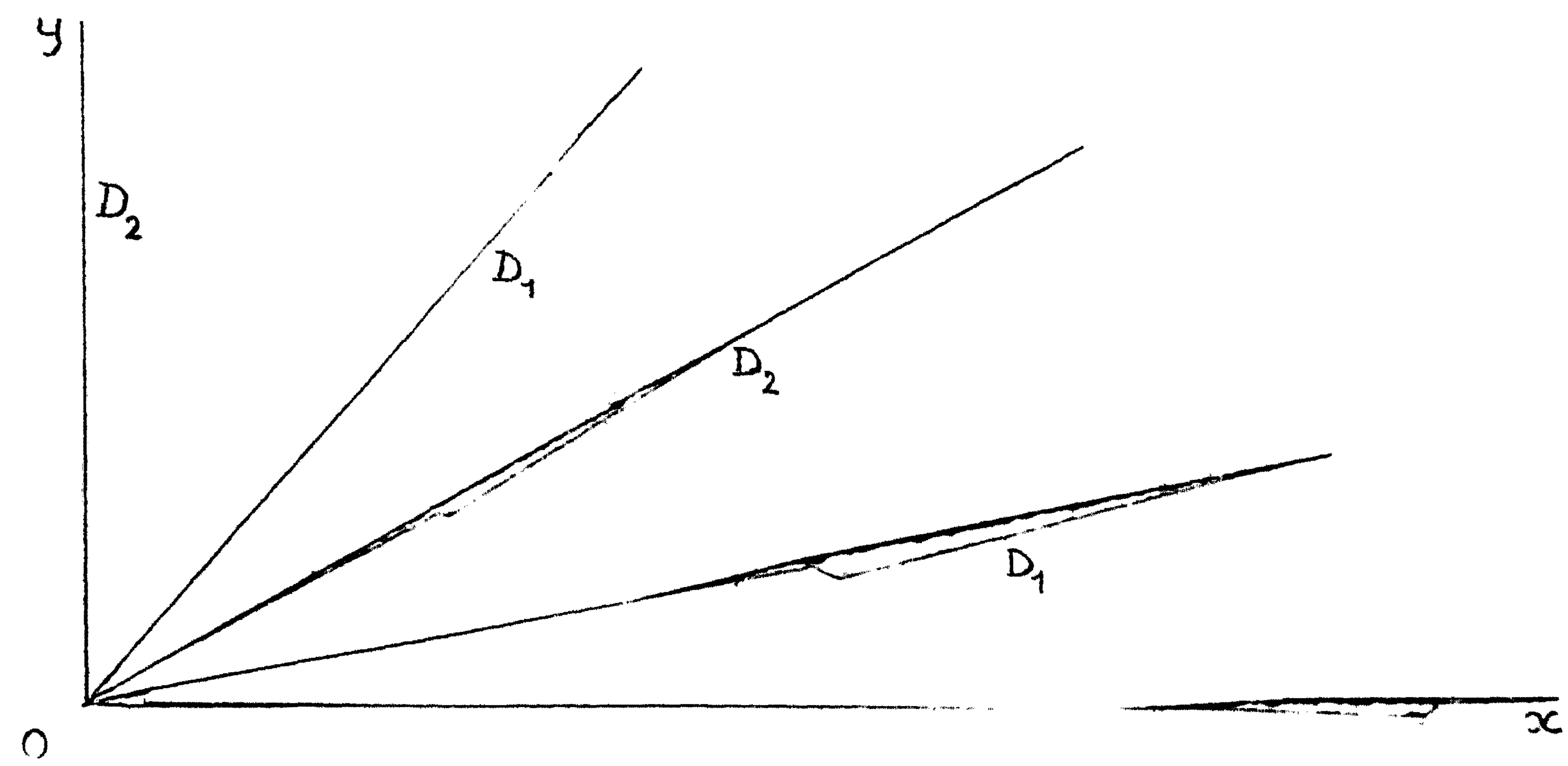


fig. 4.1

<u>Een "op het eerste gezicht" mogelijke verdeling van de toestandsruimte.</u>

In ieder van de sectoren is steeds dezelfde beslissing optimaal.

Laten wij eens nagaan voor welke punten in de toestandsruimte het onverschillig is of men eerst de beslissing D_1 neemt, gevolgd door D_2 om vervolgens optimaal verder te gaan dan wel dat men begint met D_2 en daarna de beslissing D_1 neemt alvorens optimaal verder te beslissen. De verwachte opbrengsten zullen wij aangeven met $f_{D_1 D_2}(x,y)$ resp. $f_{D_2 D_1}(x,y)$. Men kan eenvoudig nagaan, dat

$$f_{D_1 D_2}(x,y) = (1-p_1)r_1 x + (1-p_1)(1-p_2)r_2 y + (1-p_1)(1-p_2)\; f\left[(1-r_1)x,\ (1-r_2)y\right] \tag{4.26}$$

$$f_{D_2 D_1}(x,y) = (1-p_2)r_2 y + (1-p_2)(1-p_1)r_1 x + (1-p_1)(1-p_2)\; f\left[(1-r_1)x,\ (1-r_2)y\right]. \tag{4.27}$$

We stellen nu $f_{D_1 D_2}(x,y) = f_{D_2 D_1}(x,y)$,

waaruit volgt

$$(1-p_1)r_1x+(1-p_1)(1-p_2)r_2y = (1-p_2)r_2y+(1-p_1)(1-p_2)r_1x \qquad (4.29)$$

of
$$y = \frac{(1-p_1)p_2r_1}{(1-p_2)p_1r_2}\,x \quad . \qquad (4.30)$$

Uit deze relatie volgt, dat de gezochte punten liggen op een rechte $\mathcal{L}$ door de oorsprong (zie fig. 4.2).

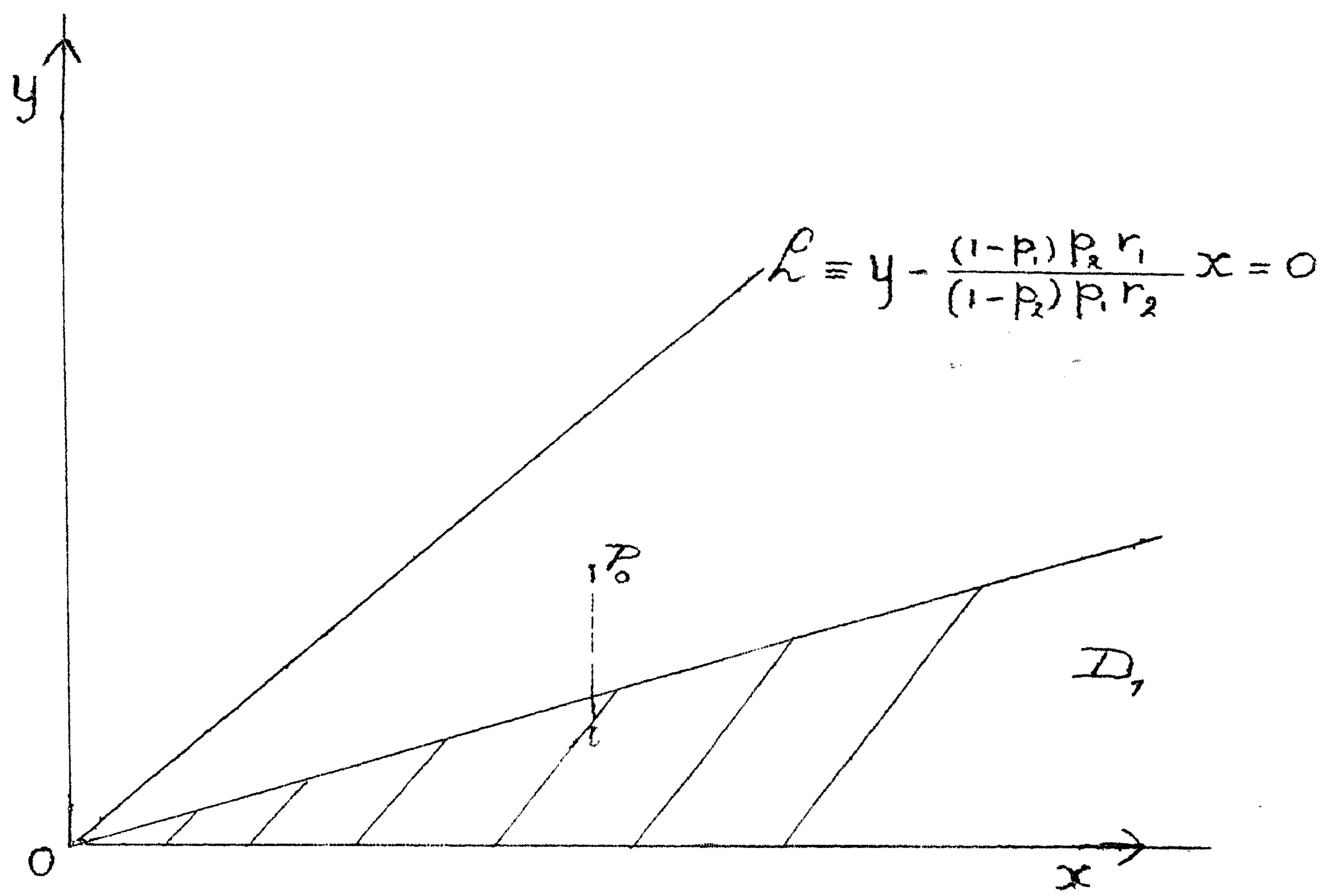


fig. 4.2.

Verder blijkt uit (4.26) en (4.27), dat voor alle punten onder $\mathcal{L}$ geldt:

$$f_{D_2D_1}(x,y) < f_{D_1D_2}(x,y) \qquad (4.31)$$

terwijl voor de punten onder $\mathcal{L}$ het omgekeerde waar is. Uit de continuïteitseigenschappen van $f(x,y)$ volgt, dat er een sector grenzend aan de x-as bestaat, waarin de beslissing D_1 te verkiezen is boven D_2. Stel nu dat er een punt P_0 bestaat beneden de lijn $\mathcal{L}$ met coördinaten (x_0,y_0) waarvoor geldt dat D_2 de optimale beslissing is. Stel vervolgens dat P_0 zodanig gelegen is dat het punt $[x_0,(1-r_2)y_0]$, voorstellende de toestand op het volgende beslissingstijdstip, ligt in de bovengenoemde sector. Uit deze eigenschappen van P_0 volgt:

$$f_{D_2 D_1}(x_0, y_0) = f(x_0, y_0) \qquad (4.32)$$

en $$f_{D_2 D_1}(x_0, y_0) < f_{D_1 D_2}(x_0, y_0). \qquad (4.33)$$

Deze beide relaties zijn echter strijdig, zodat er geen P_0 bestaat met bovengenoemde eigenschappen. Bijgevolg zijn er geen punten beneden de $\mathcal{L}$ lijn, waarvoor de beslissing D_2 optimaal is. Op dezelfde wijze kan men aantonen, dat voor de punten boven $\mathcal{L}$ de beslissing D_2 optimaal is.

<u>De optimale beslissing wordt dus gegeven door de volgende beslisregel</u>[1]

Ga vissen in V_1 als $y < \dfrac{(1-p_1)p_2r_1}{(1-p_2)p_1r_2}x$. (4.34)

Ga vissen in V_2 als $y > \dfrac{(1-p_1)p_2r_1}{(1-p_2)p_1r_2}x$.

Wij hebben tot dusver aangenomen dat de fracties r_1 en r_2 constant zijn. Zonder enig bezwaar mogen deze ook trekkingen zijn uit een verdeling. In (5.34) moet men dan r_1 vervangen door $\mathcal{E}\underline{r}_1$ en r_2 door $\mathcal{E}\underline{r}_2$.

Verder is het mogelijk, als geldt

$$(1-r_1)^a = (1-r_2)^b \qquad (4.35)$$

waarbij a en b gehele getallen zijn, de functie $f(x,y)$ te bepalen. <u>De grootheden r_1 en r_2 zijn nu constanten.</u>

Kiezen wij ter illustratie a=2 en b=1 en beschouwen wij wederom de lijn $\mathcal{L}$ van alle punten, waarvoor het onverschillig is of men eerst D_1 kiest dan wel D_2. De lijnen $\mathcal{L}_i$ worden gevormd door die punten, waarvoor geldt dat men volgens de beslisregel i maal D_1 moet kiezen terwijl de keuze voor de $(i+1)^{ste}$ keer zowel D_1 als D_2

----

1) De hierboven geschetste werkwijze kan gevolgd worden bij die problemen, welke de volgende functionaalvergelijkingen voortbrengen:

$$f(x,y) = \max \begin{cases} \int_0^1 \left\{ zx + f\left[(1-z)x, y\right] \right\} \, dG(z) \\ \int_0^1 \left\{ wy + f(x,(1-w)y) \right\} \, dH(w) \end{cases}$$

De integralen in deze functionaalvergelijking zijn Stieltjes integralen.

mag zijn. De lijnen $\mathcal{M}_j$ worden gevormd door die punten, waarvoor geldt dat men volgens de beslisregel j maal I_2 moet kiezen terwijl de keuze voor de $(j+1)^{ste}$ keer zowel I_1 als I_2 mag zijn.

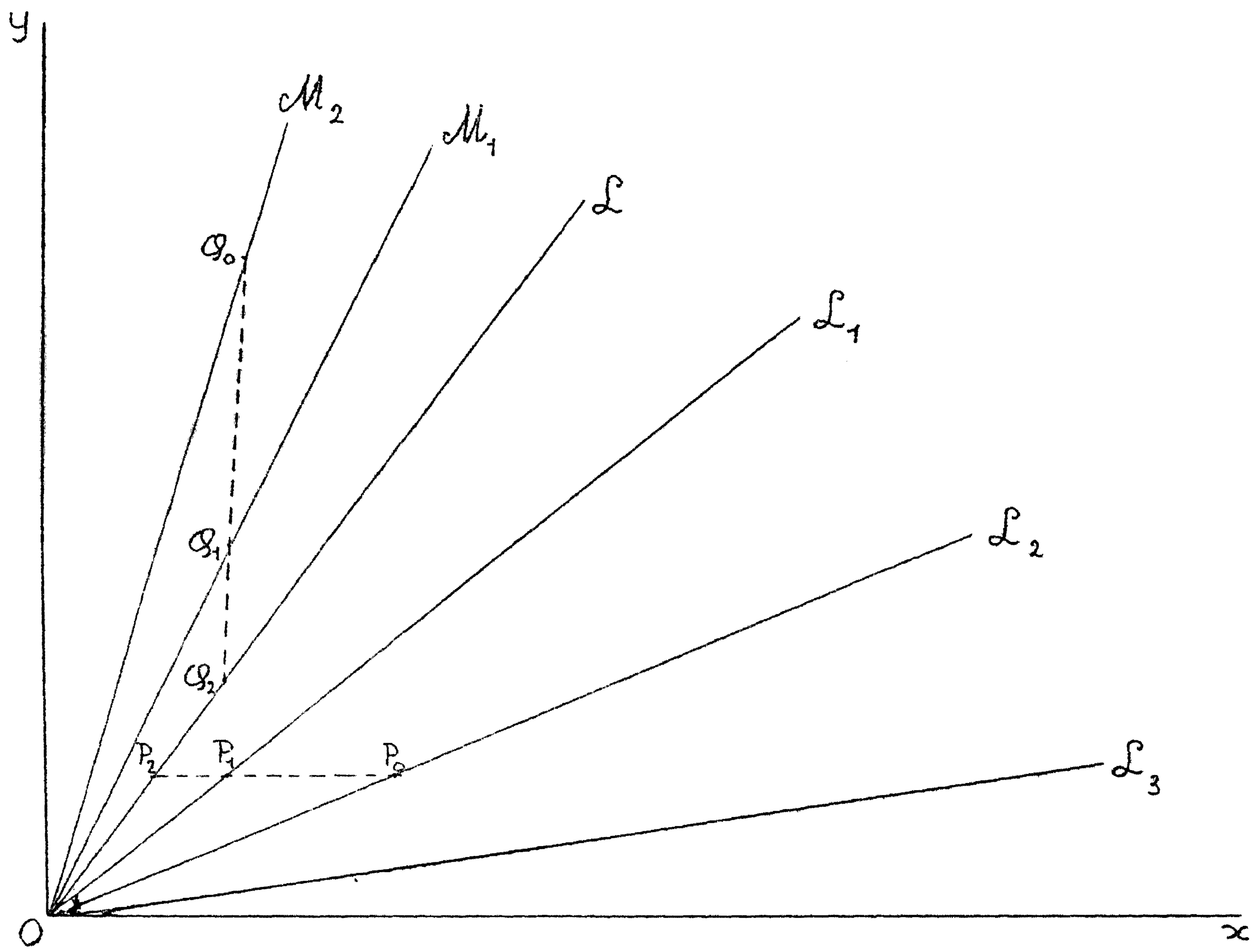


fig. 4.3

In fig. 4.3 hebben wij twee begintoestanden getekend, nl. P_0 en Q_0. De toestanden op de volgende beslissingstijdstippen worden gegeven door P_1 en P_2 resp. door Q_1 en Q_2.

Men kan ook zeggen, dat de lijn $\mathcal{L}_i$ door een beslissing getransformeerd wordt in de lijn $\mathcal{L}_{i-1}$ enz.

Men kan eenvoudig nagaan dat de lijnen $\mathcal{L}_i$ en $\mathcal{M}_j$ resp. gegeven worden door:

$$\mathcal{L}_i \equiv y - \frac{(1-p_1)p_2r_1(1-r_1)^i}{(1-p_2)p_1r_2}x = o \qquad (4.36)$$

$$\mathcal{M}_j = y - \frac{(1-p_1)p_2r_1}{(1-p_2)p_1r_2(1-r_2)^j}x = o. \qquad (4.37)$$

Voor alle punten in de sector $\mathcal{L}_1 O \mathcal{L}_2$ geldt, dat het optimaal is om tweemaal de beslissing I_1 te nemen en daarna de beslissing I_2. Wij vinden dus voor een punt (x,y) in deze sector:

$$f(x,y)=(1-p_1)r_1x+(1-p_1)^2r_1^2x+(1-p_1)^2(1-p_2)r_2y+ (1-p_1)^2(1-p_2)f\left[(1-r_1)^2x,(1-r_2)y\right]. \qquad (4.3ó)$$

Uit (4.38) en (4.35) volgt met a=2 en b=1:

$$f(x,y) = (1-p_1)r_1x+(1-p_1)^2r_1^2x+(1-p_1)^2(1-p_2)r_2y + +(1-p_1)^2(1-p_2)(1-r_1)^2f(x,y) \qquad (4.39)$$

of

$$f(\ ,y) = \frac{(1-p_1)r_1\left[1+(1-p_1)r_1\right]x+(1-p_1)^2(1-p_2)r_2y}{1-(1-p_1)^2(1-p_2)(1-r_1)^2} . \qquad (4.40)$$

De gedaante van de functie $f(x,y)$ in de sector $\mathcal{L}_2 O \mathcal{L}_3$ vindt men als volgt. Stel (x,y) is een punt in deze sector, dan is $\left[(1-r_1)x,y\right]$ een punt uit $\mathcal{L}_1 O \mathcal{L}_2$. Voor de functie $f(x,y)$ geldt dus:

$$f(x,y) = (1-p_1)r_1x+(1-p_1)f\left[(1-r_1)x,y\right] =$$
$$=(1-p_1)r_1x+(1-p_1)\frac{(1-p_1)r_1\left[1+(1-p_1)r_1\right](1-r_1)x+(1-p_1)^2(1-p_2)r_2y}{1-(1-p_1)^2(1-p_2)(1-r_1)^2}. \qquad (4.41)$$

De gedaante van de functie $f(x,y)$ in de sector $\mathcal{L} O \mathcal{L}_1$ vindt men als volgt. Stel (x,y) is een punt in deze sector, dan is $\left[\frac{x}{1-r_1}, y\right]$ een punt uit $\mathcal{L}_1 O \mathcal{L}_2$. Nu geldt

$$f\left(\frac{x}{1-r_1}, y\right) = (1-p_1)r_1\frac{x}{1-r_1} + (1-p_1)f(\ ,y) \qquad (4.42)$$

$$\text{of } f(x,y) = \frac{f\left(\frac{x}{1-r_1},y\right)-(1-p_1)r_1\frac{x}{1-r_1}}{1-p_1} =$$

$$= \frac{(1-p_1)r_1\left[1+(1-p_1)r_1\right]\frac{x}{1-r_1}+(1-p_1)^2(1-p_2)r_2y}{(1-p_1)\left[1-(1-p_1)^2(1-p_2)(1-r_1)^2\right]} - \frac{(1-p_1)r_1\frac{x}{1-r_1}}{(1-p_1)} =$$

$$\frac{r_1\left[1+(1-p_1)r_1\right]\frac{x}{1-r_1}+(1-p_1)(1-p_2)r_2y}{1-(1-p_1)^2(1-p_2)(1-r_1)^2} - r_1\frac{x}{1-r_1} . \qquad (4.43)$$

Op de hierboven geschetste werkwijze kan men voor alle sectoren de functie $f(x,y)$ bepalen.

Aangezien de sectorgrenzen gegeven worden door $\mathcal{L}_i$ en $\mathcal{M}_j$ kan men voor ieder punt (x,y) de sector bepalen. Indien men de sector kent beschikt men ook over de gedaante van de functie $f(x,y)$ in die sector en kan de functiewaarde in (x,y) worden berekend.

5. Markovian decision processes

5.1. Inleiding

In deze paragraaf zullen wij nog eens een ∞-staps-beslissingsprobleem beschouwen, waarbij men op van te voren vastgestelde tijdstippen een beslissing kan nemen. Evenals in het voorgaande zullen we hier onder een strategie χ een voorschrift verstaan dat voor elk beslissingstijdstip aangeeft welke beslissing er in de aangenomen toestand "moet" worden genomen. Een strategie voegt dus aan elk beslissingstijdstip t een functie $X_t(S)$ toe, die aangeeft welke beslissing genomen moet worden als het systeem zich op dit tijdstip in de toestand S bevindt. In het algemeen zullen deze functies voor verschillende waarden van t verschillend zijn.

Indien het systeem slechts een eindig aantal toestanden $S_1, S_2, \ldots, S_N$ kan aannemen, en het verloop van het systeem gestoord wordt volgens een zekere strategie χ, dan kan zich het geval voordoen dat het zo ontstane proces een Markov-proces is (met een eindig aantal toestanden). In zo'n geval zullen in het algemeen alle overgangswaarschijnlijkheden van het proces afhankelijk zijn van de toegepaste strategie χ; in verband hiermee zullen we de kans op een toestandsverandering van S_i naar S_j (in één stap) aangeven met $p_{ij}(\chi)$. Naar analogie hiervan zullen we de kans op een overgang van S_i naar S_j, in k stappen, aangeven met

$$p_{ij}^{(k)}(\chi).$$

Voor de laatstgenoemde overgangswhn geldt: [1)]

$$p_{ij}^{(k)}(\chi) = \sum_{s=1}^{N} p_{is}^{(k-1)}(\chi)\, p_{sj}^{(1)}(\chi). \qquad (5.1)$$

Men kan nu bewijzen, dat de limiet van

----

1) We veronderstellen dat het Markovproces enkelvoudig en stationair is.

$$\frac{1}{n}\sum_{k=1}^{n} p_{ij}^{(k)}(\chi) \qquad (5.2)$$

voor $n \to \infty$ bestaat; we schrijven voor deze limiet $q_{ij}(\chi)$ en noemen deze grootheden de invariante kansen van het betreffende Markovproces.

Stel dat $h(S_i;\chi)$ de directe opbrengst is als men in de toestand S_i een beslissing neemt volgens χ.

Als het systeem zich op het eerste beslissingstijdstip in de toestand S_i bevindt dan zal bij toepassing van de strategie χ, de verwachting van de opbrengst in de eerste m stappen gelijk zijn aan

$$h(S_i;\chi) + \sum_{j=1}^{N} h(S_j;\chi).\, p_{ij}^{(1)}(\chi) + \sum_{j=1}^{N} h(S_j;\chi)\, p_{ij}^{(2)}(\chi) +$$

$$+ \dots + \sum_{j=1}^{N} h(S_j;\chi)\, p_{ij}^{(m-1)}(\chi) =$$

$$= h(S_i;\chi) + \sum_{j=1}^{N} h(S_j;\chi).\, \sum_{k=1}^{m-1} p_{ij}^{(k)}(\chi)\ . \qquad (5.3)$$

Indien men een strategie zoekt waarvoor de opbrengst in de eerste m stappen maximaal is, dan zal men die strategie moeten bepalen waarvoor (5.3) maximaal is.

Het maximaliseren van (5.3) is gelijkwaardig met het maximaliseren van

$$\frac{h(S_i;\chi)}{m-1} + \frac{1}{m-1} \sum_{j=1}^{N} h(S_j;\chi) \sum_{k=1}^{m-1} p_{ij}^{(k)}(\chi) =$$

$$= \frac{h(S_i;\chi)}{m-1} + \sum_{j=1}^{N} h(S_j;\chi).\, \frac{\sum_{k=1}^{m-1} p_{ij}^{(k)}(\chi)}{m-1}\ . \qquad (5.4)$$

Als men m onbeperkt laat toenemen, dan gaat (5.4) over in

$$\sum_{j=1}^{N} h(S_j;\chi).\, q_{ij}(\chi). \qquad (5,5)$$

waaruit volgt dat men als <u>kriterium</u> voor de optimale strategie moet kiezen

$$H(S_i;\chi) = \sum_{j=1}^{N} h(S_j;\chi) \cdot q_{ij}(\chi) . \qquad (5.6)$$

Met behulp van (5.6) kunnen nu twee verschillende strategieën op hun merites worden onderzocht.

5.2. De „Policy improvement" methode.

In deze paragraaf zullen wij, gegeven een beslissingsprobleem, de klasse van alle strategieën beschouwen, waarvoor geldt dat zij bij toepassing van een Markovproces voortbrengen, waarin geen kring-fuiken voorkomen.
In Hoofdstuk IX hebben wij in een tweetal voorbeelden klassen bestaande uit eindig veel strategieën beschouwd, waarvoor het bovenstaande gold; verder bepaalden wij daar, uit de gegeven klassen, de optimale strategie.
Bij elk der gegevens strategieën berekenden wij de matrix der directe en daaruit die van de invariante overgangswhn om daarna de bijbehorende waarde van de kriteriumfunctie te kunnen bepalen. Lettende op de diverse waarden der kriteriumfunctie $H(S,\chi)$ kon vervolgens uit de gegeven strategieën de beste gekozen worden.
Wanneer het aantal gegeven strategieën groot is, dan zal aan deze methode het bezwaar verbonden zijn, dat er veel rekenwerk mee gemoeid is.
De toegepaste methode is dan ook zeer bewerkelijk als wij uit een grote klasse van strategieën de optimale strategie wensen te bepalen.
In deze gevallen is het daarom raadzaam gebruik te maken van de z.g.n. „policy improvement" methode.
De χ-functie $H(S_i;\chi)$, gegeven door de uitdrukking (5.6), kan worden geïnterpreteerd als de verwachting van de opbrengst te maken in een „verafgelegen" tijdsinterval tussen twee opeenvolgende equidistante beslissingstijdstippen (als S_i de begintoestand en χ de toe te passen strategie is).

Bij een ∞-stapsbeslissingsprobleem zullen, gezien vanuit de begintoestand, „bijna alle" tijdsintervallen tussen twee opeenvolgende equidistante beslissingstijdstippen ver-af liggen. Het is dus zinvol om de χ- functie $H(S_i;\chi)$ als kriterium voor de optimale strategie te kiezen.
Het is gemakkelijk in te zien dat geldt

$$H(S_i;\chi) = \sum_{j=1}^{N} p_{ij}(\chi).\ H(S_j;\chi). \qquad (5.7)$$

Beschouwen we nu een „willekeurige" strategie χ_o.
Indien niet alle waarden $H(S_i;\chi_o)$ ($i=1,2,\ldots,N$) gelijk zijn, dan is het op grond van de hierboven gegeven interpretatie van $H(S_i;\chi_o)$ zinvol om te trachten het systeem naar een toestand S_k over te brengen waarvoor geldt

$$H(S_k;\chi_o) = \max_i \ H(S_i;\chi_o). \qquad (5.8)$$

Beschouwen wij nu een situatie waarin wij de strategie χ_o zullen toepassen vanaf het tweede beslissingstijdstip terwijl we op het eerste beslissingstijdstip vrij zijn in de keuze van de beslissing X .
Nu weten wij dat de kansverdeling van de toestand op het tweede beslissingstijdstip volledig wordt bepaald door de toestand op het eerste beslissingstijdstip en de op dat tijdstip te nemen beslissing. Schrijven wij nu

$$p_{ij}\ (X) = \text{Prob}\left[\underline{S}^{(2)} = S_j \,\middle|\, S_i;\ X\right] \qquad (5.9)$$

dan zullen wij uiteraard trachten X zodanig te kiezen dat

$$\sum_{j=1}^{N} p_{ij}(X)\ \ H(S_j;\chi_o) \qquad (5.10)$$

maximaal is.
Wij gaan nu voor iedere begintoestand S_i ($i=1,\ldots,N$) de beslissing X bepalen, waarvoor geldt dat (5.10) maximaal is.
Het is niet uitgesloten dat er bij een toestand S_i meerdere optimale waarden X bestaan.
We definiëren nu $D_O(S_i)$ als de verzameling van alle beslissingen X waarvoor geldt

$$\sum_{j=1}^{N} p_{ij}(X) \quad H(S_j, \chi_0) = \max_{X'} \sum_{j=1}^{N} p_{ij}(X') \, H(S_j; \chi_0)$$

in formule

$$D_0(S_i) = \left\{ X \sum_{j=1}^{N} p_{ij}(X) \quad H(S_j, \chi_0) = \max_{X'} \sum_{j=1}^{N} p_{ij}(X') H(S_j; \chi_0) \right. \tag{5.11}$$

Breiden wij nu de probleemstelling een weinig uit.
Stel dat de beslisser alleen dan het systeem mag beheren als hij op elk beslissingstijdstip een „pachtsom" betaalt gelijk aan $H(S_j; \chi_0)$, waarbij S_j de toestand van het systeem aangeeft op het betreffende beslissingstijdstip.
Door deze keuze van de pachtsom zal de verwachting van de winst voor „verafgelegen" tijdsintervallen gelijk aan nul zijn.
Beschouwen wij wederom de situatie, waarin de strategie χ_0 zal worden toegepast vanaf het tweede beslissingstijdstip en waarin de beslisser de eerste beslissing willekeurig uit $D_0(S_i)$ mag kiezen.
Als S_i de begintoestand is dan wordt de verwachting van de(netto) winst voor de beslisser voor de eerste M tijdsintervallen gegeven door:

$$g_M(S_i; X; \chi_0) = h(S_i; X) - H(S_i; \chi_0) +$$

$$\sum_{m=0}^{M-2} \sum_{j=1}^{N} \sum_{k=1}^{N} (p_{ij}(X) \, p_{jk}^{(m)}(\chi_0) \left[h(S_k; \chi_0) - H(S_k; \chi_0) \right]. \tag{5.12}$$

De te behalen winst in een onbegrensd aantal perioden wordt nu gegeven door

$$g(S_i; X; \chi_0) = \lim_{M \to \infty} g_M(S_i; X; \chi_0) \;{}^{1)}$$

1) Deze limiet bestaat niet als het Markovproces met overgangswhn. $p_{ij}(\chi_0)$ kringfuiken bevat. Dit bezwaar kan echter ondervangen worden. Wij zullen deze gevallen evenwel buiten beschouwing laten.

Wij kiezen nu bij iedere toestand S_i een beslissing $X \in D_0(S_i)$, waarvoor de X-functie $g(S_i;X;\chi_0)$ maximaal is. Indien dit maximum ook wordt bereikt voor de beslissing X die we volgens χ_0 zouden moeten nemen dan kiezen we deze laatste; is dit niet het geval dan zijn we vrij in de keuze van de X waarde als (5.13) voor deze X maar maximaal is.
Op deze wijze verkrijgen wij bij iedere toestand S_i een beslissing X; m.a.w. we vinden een nieuwe strategie. Deze strategie zullen we aangeven met χ_1. De procedure die wij hier-boven hebben aangegeven herhalen wij met χ_1 i.p.v. χ_0 2). Op deze wijze vinden we een strategie χ_2 etc. Men kan bewijzen dat de rij strategieën $\chi_1, \chi_2, \ldots.$ convergeert naar de optimale strategie χ^*.

---

2) We nemen aan dat ook de strategie χ_1 tot de in het begin van deze paragraaf genoemde klasse behoort.