stichting
mathematisch
centrum

AFDELING MATHEMATISCHE BESLISKUNDE BC 19/77 SEPTEMBER

O.J. VRIEZE

MODELVORMING MET STOCHASTISCHE SPELEN

2e boerhaavestraat 49 amsterdam

Printed at the Mathematical Centre, 49, 2e Boerhaavestraat, Amsterdam.

The Mathematical Centre, founded the 11-th of February 1946, is a non-profit institution aiming at the promotion of pure mathematics and its applications. It is sponsored by the Netherlands Government through the Netherlands Organization for the Advancement of Pure Research (Z.W.O).

AMS(MOS) subject classification scheme (1970): 90D15, 90D45

Modelvorming met stochastische spelen

door

O.J. Vrieze

SAMENVATTING

Dit rapport is samengesteld naar aanleiding van de werkweek "Modelvorming met Speltheorie", georganiseerd door de afdeling Mathematische Besliskunde van het Mathematisch Centrum, welke is gehouden in september 1977 en alwaar de auteur een lezing heeft verzorgd met betrekking tot het onderwerp "Stochastische spelen en haar toepassingen".

In hoofdstuk 1 wordt een inleiding gegeven in de theorie der twee-persoons stochastische spelen. Zowel nulsom als niet-nulsom spelen komen aan de orde. De beschouwde modellen zijn eindig en stoppend.
Hoofdstuk 2 behandelt toepassingen van de theorie der twee-persoons nulsom spelen op zoekmodellen. Een aantal voorbeelden worden uitgewerkt.
In hoofdstuk 3 wordt de theorie der twee-persoons niet-nulsom spelen toegepast op duopoly modellen. Het accent ligt hierbij op berekenbare evenwichtspunten.

TREFWOORDEN: *stochastische spelen, toepassingen van stochastische spelen, zoekmodellen, duopoly modellen.*

INHOUD

HOOFDSTUK 1

Theorie der stochastische spelen.

1.1. Inleidende begrippen

In dit hoofdstuk wordt de theorie van de eindige, zowel nulsom als niet-nulsom, twee-persoons stochastische spelen uiteengezet.
Het woordje "eindige" slaat op het eindig zijn van zowel de toestandsruimte als de aktieruimten van de beide spelers in iedere toestand.
We veronderstellen, dat de lezer bekend is met de voornaamste resultaten op het gebied van de matrix spelen en bimatrix spelen.
De theorie van de nulsom spelen die we zullen behandelen is gebaseerd op SHAPLEY [14]. De theorie van de niet-nulsom spelen die we zullen behandelen is gebaseerd op ROGERS [12] en VRIEZE [20].

Een stochastisch spel wordt gedefinieerd door vijf objecten:
Een eindige verzameling I van spelers.
Een verzameling S van toestanden.
Een verzameling A van akties.
Een kansmaat q.
Een opbrengstfunctie r_i voor iedere speler $i \in I$.

Gedragsmatig is het stochastische spel een rij $\gamma_1, \gamma_2, \ldots$ van niet-coöperatieve niet-nulsom spelen gespeeld door de leden van I. S indiceert dan de verzameling $\{\Gamma_s \mid s \in S\}$ waarvan γ_t voor alle $t = 1,2,\ldots$ een element is. Voor iedere toestand $s \in S$ is Γ_s dus een niet-coöperatief spel met evenveel spelers als het aantal leden van I. De aanduiding "de toestand van het systeem is s" komt dus overeen met "spel Γ_s wordt gespeeld".
De volgorde van de te spelen spellen wordt op stochastische wijze bepaald door de kansmaat q en wel op de volgende wijze:

Als op tijdstip t de toestand van het systeem s_t is (dus $\gamma_t = \Gamma_{s_t}$) en als de spelers in toestand s_t repsektievelijk akties $a_n(s_t) \in A$, alle $n \in I$, kiezen, dan specificeert q een kansmaat $p(\cdot|s_t, a(s_t))$, op de Borel deelverzamelingen van S, met $\bar{a}(s_t) = (a_1(s_t), \ldots, a_M(s_t))$ en M het aantal leden van I. Is H een Borel deelverzameling, dan is $p(H|s_t, \bar{a}(s_t))$ de kans, dat de volgende toestand een element van H is. In dit rapport behandelen we vrijwel uitsluitend systemen waarbij de toestandsruimte uit een eindig aantal elementen bestaat, bijvoorbeeld genummerd 1 t/m N. Dan is $p(\cdot|s_t,\bar{a}(s_t))$ een rij $p(1|s_t,\bar{a}(s_t)),p(2|s_t,\bar{a}(s_t)),\ldots,p(N|s_t,\bar{a}(s_t))$ met $p(s|s_t,\bar{a}(s_t)) \geq 0, \forall s \in \{1,\ldots,N\}$ en $\sum_{s=1}^{N} (s|s_t,\bar{a}(s_t)) \leq 1$. De kansen $p(s|s_t,\bar{a}(s_t))$ worden de overgangskansen genoemd. Als $\sum_{s=1}^{N} p(s|s_t,\bar{a}(s_t)) < 1$, dan is $1 - \sum_{s=1}^{N} p(s|s_t,\bar{a}(s_t))$ de kans dat het spel stopt na toestand s_t als de gezamenlijke akties $\bar{a}(s_t)$ gekozen worden.

Als in toestand s_t op tijdstip t de spelers respektievelijk $a_n(s_t) \in A$, alle $n \in I$, kiezen, dan is er een direkte opbrengst voor de spelers, die voor speler n gelijk is aan $r_n(s,\bar{a}(s_t))$.
In het algemeen is de verzameling van mogelijke akties van een bepaalde speler in een bepaalde toestand niet de gehele verzameling A maar een deelverzameling hiervan, welke afhankelijk is van de toestand.

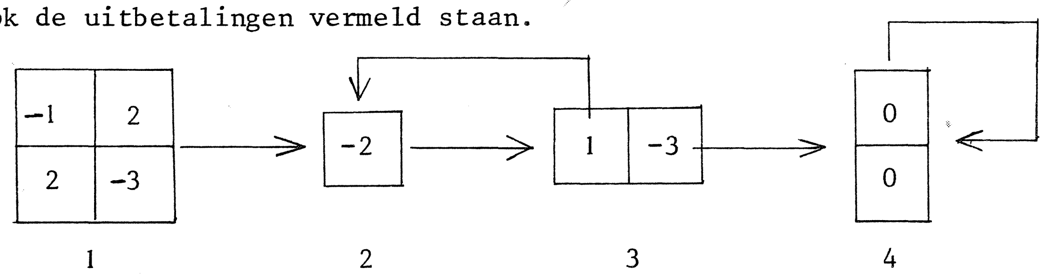
In dit rapport leggen we de volgende restricties op op de spelparameters.
R1: I bestaat uit twee elementen
R2: S bevat een eindig aantal elementen
R3: A bevat een eindig aantal elementen.

<u>DEFINITIE</u>. Gelden voor een stochastisch spel de restricties R1, R2 en R3, dan spreken we van een eindig twee-persoons stochastisch spel.

Alle stochastische spelen die in dit hoofdstuk aan de orde komen zijn twee-persoons. In het vervolg spreken we over een stochastisch spel als we een eindig twee-persoons stochastisch spel bedoelen.

<u>VOORBEELD</u>. Vier toestanden genummerd 1,2,3 en 4. In toestand 1 hebben beide spelers 2 akties ter beschikking. In toestand 2 hebben beide spelers 1 aktie ter beschikking. In toestand 3 heeft speler 1 één aktie en speler 2 twee akties ter beschikking. In toestand 4 heeft speler 1 twee akties en speler 2 één aktie ter beschikking. In toestand 4 heeft speler 1 twee akties en speler 2 één aktie ter beschikking.
Vanuit toestand 1 springt het spel onder ieder paar akties naar toestand 2.
Vanuit toestand 2 springt het spel onder ieder paar akties naar toestand 3.
Vanuit toestand 3 springt het spel onder de eerste aktie van speler 2 naar toestand 2 en onder de tweede aktie van speler 2 naar toestand 4. In toestand 4 springt het spel onder de beide akties van speler 1 naar toestand 4.

Schematisch valt deze situatie weer te geven m.b.v. de volgende figuur waarin ook de uitbetalingen vermeld staan.



Dit is een voorbeeld van een nulsom spel. De getallen in de hokjes zijn de uitbetalingen aan speler 1 behorende bij de overeenkomstige akties. De uitbetalingen aan speler 2 zijn het tegengestelde van deze waarden. Zoals gebruikelijk voor matrix spelen komt het kiezen van een zuivere aktie voor speler 1 overeen met het kiezen van een rij van de matrix en voor speler 2 met het kiezen van een kolom. In dit voorbeeld zijn alle overgangen deterministisch. Dit is in het algemeen uiteraard niet zo.
Als het spel niet start in toestand 1, zal toestand 1 nooit bereikt worden. Vanuit iedere toestand kan speler 2 ervoor zorgen, dat het spel in toestand 4 komt, waarna dan spel 4 herhaald gespeeld wordt. Speler 2 zal er echter de voorkeur aan geven het spel heen en weer te laten springen tussen de toestanden 2 en 3.

Laat $A_n(s) \subset A$ de zuivere aktieruimte zijn voor speler n in toestand $s \in S$, $n = 1,2$. Laat $\Pi_n(s)$ de verzameling van gemengde akties voor speler n in toestand s zijn. Laat $q_0(s,a_1(s),a_2(s)) = 1 - \sum_{s' \in S} p(s' | s,a_1(s),a_2(s))$.

<u>DEFINITIE</u>. Een stochastisch spel wordt nulsom genoemd als voor alle $s \in S$, $a_1(s) \in A_1(s)$ en $a_2(s) \in A_2(s)$ geldt $r_1(s,a_1(s),a_2(s)) = - r_2(s,a_1(s),a_2(s))$.

<u>DEFINITIE</u>. Een stochastisch spel wordt stoppend genoemd als er een getal $\varepsilon > 0$ bestaat z.d.d. voor alle $s \in S, a_1(s) \in A_1(s)$ en $a_2(s) \in A_2(s)$ geldt $q_0(s,a_1(s),a_2(s)) \geq \varepsilon$.

<u>DEFINITIE</u>. Een stochastisch spel wordt niet-stoppend genoemd als voor alle $s \in S$, $a_1(s) \in A_1(s)$ en $a_2(s) \in A_2(s)$ geldt $q_0(s,a_1(s),a_2(s)) = 0$.

Merk op, dat er spelen zijn, die noch stoppend noch niet-stoppend zijn. Het spel in bovenstaand voorbeeld is niet-stoppend. We gaan nu het begrip strategie voor een speler definiëren; hierbij hebben we het begrip historie van een spel nodig.

DEFINITIE. De historie van een stochastisch spel op tijdstip t (notatie h_t) is $h_t = (s_0,\bar{a}(s_0),s_1,\bar{a}(s_1),\dots,s_{t-1},\bar{a}(s_{t-1}),s_t)$ als s_τ respektievelijk $\bar{a}(s_\tau)$ de toestand respektievelijk de akties van de spelers op tijdstip τ was, $\tau = 0,1,\dots,t$.

DEFINITIE. Een gedragsstrategie voor speler n (notatie π_n^G) is een voorschrift, dat aan ieder tijdstip t en iedere historie h_t een element van $\Pi_n(s_t)$ toevoegt (notatie $\pi_n^G(t,h_t)$).

Laat Π_n^G de verzameling van gedragsstrategieën voor speler n zijn. We zouden voor speler n nog een ruimere klasse van strategieën kunnen definiëren dan de klasse Π_n^G. Echter AUMANN [1] heeft bewezen, dat, wat stochastische spelen betreft, we ons zonder aan de algemeenheid afbreuk te doen, kunnen beperken tot de klasse Π_n^G. Een element van Π_n^G wordt in het vervolg kortweg een strategie genoemd.

DEFINITIE. Een strategie wordt semi-markov genoemd als deze voor iedere t wat de historie betreft, alleen afhankelijk is van s_0 en s_t.

DEFINITIE. Een strategie wordt markov of geheugenloos genoemd als deze voor iedere t wat de historie betreft alleen afhankelijk is van s_t.

DEFINITIE. Een strategie wordt stationair genoemd als deze voor iedere t onafhankelijk is van t en wat de historie betreft alleen afhankelijk is van s_t.

Een markov-strategie respektievelijk stationaire strategie voor speler n noteren we als π_n^M respektievelijk π_n. De klasse van markov-strategieën respektievelijk stationaire strategieën noteren we als Π_n^M respektievelijk Π_n. Uiteraard geldt $\Pi_n \subset \Pi_n^M \subset \Pi_n^G$.
Een paar markov-strategieën π_1^M en π_2^M voor de beide spelers legt een in het algemeen niet-stationair discreet markov-proces op de toestanden van het stochastische spel vast. Eenvoudig is namelijk in te zien, dat de toestand op tijdstip t uitsluitend bepaald wordt door de toestand en de akties van

de beide spelers op tijdstip t-1 en door de gegeven overgangskansen en aangezien de akties voor markov-strategieën uitsluitend afhankelijk zijn van t-1 en s_{t-1} volgt, dat we te maken hebben met een Markov-proces. Een paar stationaire strategieën π_1 en π_2 voor de beide spelers legt een stationair discreet Markov-proces op de toestanden vast. Stationair omdat de keuze van de akties nu alleen nog maar afhangt van de toestand en niet van het tijdstip en daardoor eveneens de overgangskansen.

In het vervolg geven we het aantal toestanden van een stochastisch spel aan met N en de toestanden achten we genummers 1 t/m N. Toestanden worden aangeduid met k en ℓ, akties met i en j. Laat in toestand k m_{kn} het aantal zuivere akties voor speler n zijn. Een stationaire strategie voor speler n is dan $\pi_n = (\pi_n^1,\ldots,\pi_n^N)$, met

$$\pi_n^k = (\pi_n^k(1),\ldots,\pi_n^k(m_{kn})) \text{ en } \pi_n^k(i) \geq 0,\ \sum_{i=1}^{m_{kn}} \pi_n^k(i) = 1.$$

Laat voor een paar stationaire strategieën π_1 en π_2 $P_{\pi_1\pi_2}$ de $N \times N$ matrix zijn waarvan het $(k,\ell)^e$ element gelijk is aan:

$$\sum_{i=1}^{m_{k1}} \sum_{j=1}^{m_{k2}} \pi_1^k(i)\cdot\pi_2^k(j)\cdot p(\ell|k,i,j).$$

De waarde van het $(k,\ell)^e$ element is dus de kans, dat de volgende toestand ℓ is als in toestand k de spelers gemengde akties π_1^k en π_2^k kiezen (notatie $p(\ell|k,\pi_1,\pi_2)$).

Laat $P_{\pi_1\pi_2}^t$ het t-voudig produkt $P_{\pi_1\pi_2} \times P_{\pi_1\pi_2} \times\ldots\times P_{\pi_1\pi_2}$ zijn.

Het valt eenvoudig na te gaan, dat het $(k,\ell)^e$ element van $P_{\pi_1\pi_2}^t$ de kans is, dat, indien de beide spelers de stationaire strategieën π_1 en π_2 spelen en indien de starttoestand k is, het systeem zich na t stappen in toestand ℓ bevindt (notatie $p^t(\ell|k,\pi_1,\pi_2)$).

Voor een paar stationaire strategieën π_1 en π_2 kunnen we nu de verwachte uitbetaling op tijdstip t voor de beide spelers bepalen. We noteren deze verwachte uitbetaling voor speler n als $V_n^t(k_1\pi_1,\pi_2)$ als k de begintoestand van het spel is.

Laat $\bar{V}_n^t(\pi_1,\pi_2) = (V_n^t(1,\pi_1,\pi_2),\ldots,V_n^t(N,\pi_1,\pi_2))$.

Laat $r_n(k,\pi_1,\pi_2) = \sum_{i=1}^{m_{k1}} \sum_{j=1}^{m_{k2}} \pi_1^k(i)\cdot\pi_2^k(j)\cdot r_n(k,i,j)$.

$r_n(k,\pi_1,\pi_2)$ is dus de verwachte uitbetaling aan speler n in toestand k als de spelers π_1 en π_2 spelen.
Laat $\bar{r}_n(\pi_1,\pi_2) = (r_n(1,\pi_1,\pi_2),\ldots,r_n(N,\pi_1,\pi_2))$.
Uit het bovenstaande volgt nu:

$$\bar{V}_n^t(\pi_1,\pi_2) = P^t_{\pi_1\pi_2}\ \bar{r}_n(\pi_1,\pi_2).$$

Hierbij is $P^0_{\pi_1\pi_2}$ gedefinieerd als de eenheidsmatrix.

We kunnen nu aan een paar strategieën π_1 en π_2 op drie manieren een getal toevoegen, dat een indicatie geeft van de geschiktheid van deze strategieën voor de spelers.

1. Het verdisconteerde opbrengstenmodel.
   Er is gespecificeerd een verdisconderingsfactor β, $0 \le \beta < 1$ en de opbrengst op tijdstip t wordt vermenigvuldigd met β^t om de waarde van dit bedrag op tijdstip 0 te bepalen.
   Laat

$$\bar{W}_n^t(\pi_1,\pi_2) = \sum_{\tau=0}^{t} \beta^\tau\ \bar{V}_n^\tau(\pi_1,\pi_2)$$

   en

$$\bar{W}_n(\pi_1,\pi_2) = \lim_{t\to\infty} \bar{W}_n^t(\pi_1,\pi_2) = \sum_{\tau=0}^{\infty} \beta^\tau\ \bar{V}_n^\tau(\pi_1,\pi_2).$$

   De k^e komponent van $\bar{W}_n^t(\pi_1,\pi_2)$ respektievelijk $\bar{W}_n(\pi_1,\pi_2)$ is dan bij begintoestand k gelijk aan de verwachte verdisconteerde opbrengst tot en met tijdstip t respektievelijk de totale verwachte verdisconteerde opbrengst voor speler n.
   Laat $M_n = \max_{k,i,j}|r_n(k,i,j)|$ en laat de norm van een N-vektor gedefinieerd zijn als $\|\bar{x}\| = \max_k|x(k)|$, dan geldt:

$$\|\bar{W}_n(\pi_1,\pi_2)\| \le \sum_{\tau=0}^{\infty} \beta^\tau \|\ \bar{V}\ (\pi_1,\pi_2)\| \le M_n \sum_{\tau=0}^{\infty} \beta^\tau\ P^\tau_{\pi_1\pi_2} = \frac{M_n}{1-\beta}.$$

2. Het totale opbrengstenmodel.

De verschillende opbrengsten op de verschillende tijdstippen worden bij elkaar opgeteld.

Laat $\bar{X}_n^t(\pi_1,\pi_2) = \sum_{\tau=0}^{t} \bar{V}_n^t(\pi_1,\pi_2)$ en $\bar{X}_n(\pi_1,\pi_2) = \lim_{t\to\infty} \bar{X}_n(\pi_1,\pi_2)$, indien deze limiet bestaat.

De k^e komponent van $\bar{X}_n^t(\pi_1,\pi_2)$ respektievelijk $\bar{X}_n(\pi_1,\pi_2)$ is dan bij begintoestand k gelijk aan de verwachte opbrengst tot en met tijdstip t, respektievelijk de totale verwachte opbrengst voor speler n. Er zijn voorwaarden aan te geven waaronder $\bar{X}_n(\pi_1,\pi_2)$ bestaat. Bijvoorbeeld $\bar{X}_n(\pi_1,\pi_2)$ bestaat als we te maken hebben met een stoppend spel.

Laat $q_0 = \min_{k,i,j} q_0(k,i,j)$, dus $q_0 > 0$ is de minimale stopkans die op kan treden, dan geldt:

$$\|\bar{X}_n^t(\pi_1,\pi_2)\| \le M_n \sum_{\tau=0}^{t} P_{\pi_1\pi_2}^{\tau} \le M_n \sum_{\tau=0}^{t} (1-q_0)^{\tau}.$$

Hieruit blijkt, dat de limiet bestaat voor $t \to \infty$ en dat $\|\bar{X}_n(\pi_1,\pi_2)\| \le \frac{M_n}{q_0}$.

3. Het gemiddelde opbrengsten model.

We kijken in dit model naar de gemiddelde opbrengsten per stap.

Laat $\bar{Y}_n^t(\pi_1,\pi_2) = \frac{1}{t+1} \sum_{\tau=0}^{t} \bar{V}_n^{\tau}(\pi_1,\pi_2)$ en $\bar{Y}_n(\pi_1,\pi_2) = \lim_{t\to\infty} \bar{Y}_n^t(\pi_1,\pi_2)$, indien deze limiet bestaat.

De k^e komponent van $\bar{Y}_n^t(\pi_1,\pi_2)$ respektievelijk $\bar{Y}_n(\pi_1,\pi_2)$ is dan bij begintoestand k gelijk aan de verwachte gemiddelde opbrengst per stap tot en met tijdstip t respektievelijk de verwachte gemiddelde opbrengst per stap in het oneindige horizon spel voor speler n.

Bovenstaande definities van $\bar{W}_n^t$, $\bar{W}_n$, $\bar{X}_n^t$, $\bar{X}_n$, $\bar{Y}_n^t$ en $\bar{Y}_n$ zijn voor de overzichtelijkheid gegeven voor een paar stationaire strategieën. Echter voor ieder paar gedragsstrategieën π_1^G en π_2^G kunnen overeenkomstige vektoren gedefinieerd worden, waarvan de komponenten dezelfde betekenis hebben als in $\bar{W}_n^t$, etc. Daar we ons in dit rapport hoofdzakelijk richten op stationaire strategieën wordt dit hier achterwege gelaten.

De volgende definities hebben betrekking op een spel, waarbij de opbrengsten verdisconteerd worden. Voor het totale en gemiddelde opbrengstenmodel kunnen analoge definities opgesteld worden.

DEFINITIE. Een paar strategieën (π_1^*,π_2^*) vormt voor een niet-nulsom stochastisch spel een evenwichtspunt binnen een klasse $Q_1 \times Q_2$ van strategieën als geldt:

$$\bar{W}_1(\pi_1^G,\pi_2^*) \le \bar{W}_1(\pi_1^*,\pi_2^*), \qquad \forall \pi_1^G \in Q_1$$

$$\bar{W}_2(\pi_1^*,\pi_2^G) \le \bar{W}_2(\pi_1^*,\pi_2^*), \qquad \forall \pi_2^G \in Q_2.$$

Laat voor een nulsom stochastisch spel $\bar{W}(\pi_1,\pi_2)$ de totale verwachte verdisconteerde uitbetaling aan speler 1 bij de strategieën π_1 en π_2. De uitbetaling aan speler 2 is dan $-\bar{W}(\pi_1,\pi_2)$.

DEFINITIE. Een nulsom stochastisch spel heet strikt bepaald binnen de klasse $Q_1 \times Q_2$ van strategieën als geldt:

$$\sup_{\pi_1^G \in Q_1} \inf_{\pi_2^G \in Q_2} \bar{W}(\pi_1^G,\pi_2^G) = \inf_{\pi_2^G \in Q_2} \sup_{\pi_1^G \in Q_1} \bar{W}(\pi_1^G,\pi_2^G) = \bar{W}^*.$$

De vektor $\bar{W}^*$ wordt de waardevektor van het spel genoemd. Bestaan er strategieën π_1^* en π_2^*, z.d.d.

$$\sup_{\pi_1 \in Q_1} \bar{W}(\pi_1^G,\pi_2^*) = \bar{W}^* = \inf_{\pi_2 \in Q_2} \bar{W}(\pi_1^*,\pi_2^G) \text{ dan worden}$$

π_1^* en π_2^* optimale strategieën voor respektievelijk speler 1 en speler 2 genoemd.

Merk op, dat het bestaan van optimale strategieën in het nulsomspel het analogon is van het bestaan van evenwichtspunten in het niet-nulsomspel.

Alvorens over te gaan naar sektie 1.2., waarin we nader ingaan op het al of niet bestaan van de waarde en optimale strategieën voor nulsomspelen, vermelden we eerst nog enkele resultaten uit de theorie der matrixspelen.

In het volgende wordt met G een $m_1 \times m_2$ matrix spel bedoeld.
Laat $X_n = \{\bar{x}_n | \bar{x}_n = (x_n(1),\ldots,x_n(m_n)),\ x_n(i) \ge 0,\ i = 1,\ldots,m_n;\ \sum_{i=1}^{m_n} x_n(i) = 1\}$, $n = 1,2$. X_n is de verzameling van gemengde akties voor speler n in het matrix spel. Als de spelers respektievelijk $\bar{x}_1 \in X_1$ en $\bar{x}_2 \in X_2$ spelen,

dan is de verwachte opbrengst gelijk aan $\bar{x}_1^T \cdot G \cdot \bar{x}_2$. De waarde van een matrix spel G noteren we als Val G.

STELLING 1.1.1. *Voor ieder matrix spel geldt:*

$$\max_{\bar{x}_1 \in X_1} \min_{\bar{x}_2 \in X_2} x_1^T G x_2 = \min_{\bar{x}_2 \in X_2} \max_{\bar{x}_1 \in X_2} \bar{x}_1^T G \bar{x}_2 = \text{Val } G.$$

Een paar akties $\bar{x}_1^*$ *en* $\bar{x}_2^*$ *zijn dan en slechts dan optimaal voor respektievelijk speler 1 en speler 2 als geldt* $\bar{x}_1^T G \bar{x}_2^* \le \bar{x}_1^{*T} G \bar{x}_2^* \le \bar{x}_1^{*T} G \bar{x}_2$, *voor alle* $\bar{x}_1$ *en* $\bar{x}_2$. *Tevens is dan* $\text{Val } G = \bar{x}_1^{*T} G \bar{x}_2^*$.

Deze stelling bewijzen we niet. Het bestaan van een waarde voor ieder matrix spel is het eerst bewezen door J. VON NEUMANN [10]. Sindsdien zijn nog vele bewijzen hiervan in de literatuur verschenen.

LEMMA 1.1.2. *Zijn* G_1 *en* G_2 *twee* $m_1 \times m_2$ *spelmatrices en geldt* $g_{1ij} \le g_{2ij} + k$ *voor alle* i *en* j, *dan is* $\text{Val } G_1 \le \text{Val } G_2 + k$.

BEWIJS. Laat $\bar{x}_2^*$ een optimale strategie voor speler 2 in spel G_2 zijn, dan voor $i = 1,\dots,m_1$:

$$\sum_{j=1}^{m_2} x_2^*(j) \cdot g_{1ij} \le \sum_{j=1}^{m_2} x_2^*(j) \cdot g_{2ij} + k \le \text{Val } G_2 + k.$$

De strategie $\bar{x}_2^*$ garandeert speler 2 voor spel G_1 tegen iedere zuivere strategie van speler 1 een verwachting van de uitbetaling aan speler 1 van hoogstens $\text{Val } G_2 + k$, m.a.w. $\text{Val } G_1 \le \text{Val } G_2 + k$. □

LEMMA 1.1.3. *Zijn* G_1 *en* G_2 *twee* $m_1 \times m_2$ *spelmatrices, dan geldt:*

$$\min_{i,j} |g_{1ij} - g_{2ij}| \le |\text{Val } G_1 - \text{Val } G_2| \le \max_{i,j} |g_{1ij} - g_{2ij}|.$$

BEWIJS. Er geldt $g_{1ij} \le g_{2ij} + \max_{i,j} |g_{1ij} - g_{2ij}|$, alle i en j.
Uit lemma 1.1.2. volgt nu:

$$\text{Val } G_1 - \text{Val } G_2 \le \max_{i,j} |g_{1ij} - g_{2ij}|.$$

Op grond van symmetrie-overwegingen geldt ook

$$\text{Val } G_2 - \text{Val } G_1 \le \max_{i,j} |g_{1ij} - g_{2ij}|,$$

waarmee het rechter gedeelte van de bewering bewezen is. Op dezelfde manier vinden we met behulp van $g_{1ij} \le g_{2ij} - \min_{i,j} |g_{1ij} - g_{2ij}|$ het linker gedeelte van het lemma. □

1.2. Twee-persoons nulsom stochastische spelen

In deze sektie worden eindige twee-persoons nulsom stochastische spelen behandeld, waarbij de overgangskansen zodanig zijn, dat het spel stoppend is. SHAPLEY [14] is de eerste geweest, die deze spelen geanalyseerd heeft, vandaar, dat we ze ook wel stochastische spelen van Shapley noemen. We gebruiken dezelfde notaties als in de vorige sektie.
De situatie in toestand k kan symbolisch worden weergegeven d.m.v. een $(m_{k1} \times m_{k2})$-matrix G_k, waarbij het $(i,j)^e$ element gelijk is aan $r(k,i,j) + \sum_{\ell=1}^{N} p(\ell|k,i,j)G_\ell$. We beschouwen de stoppende spelen.
Laat

$$q_0 = \min_{k,i,j} q_0(k,i,j) = \min_{k,i,j} (1 - \sum_{=1}^{N} p(\ell|k,i,j)) > 0.$$

Als criterium nemen we het totale opbrengsten model. Zoals uit een opmerking in sektie 1.1. blijkt, bestaat voor een stoppend spel onder ieder paar stationaire strategieën de totale verwachte uitbetaling en is de absolute waarde hiervan begrensd door $\frac{M}{q_0}$ met $M = \max_{k,i,j} |r(k,i,j)|$.
Op dezelfde manier kan men ook laten zien, dat onder ieder paar strategieën de totale verwachte uitbetaling eveneens begrensd is door $\frac{M}{q_0}$.
Voor het oplossen van deze spelen beschouwen we eerst eindigstapsspelen.

DEFINITIE. Een t-staps nulsom stochastisch spel met nabetalingsvektor $\bar{R} = (R(1),\dots,R(N))$ is een stochastisch spel, dat stopt na t stappen (als het daarvoor al niet gestopt is) met een nabetaling van R(k) aan speler 1 en − R(k) aan speler 2 als k de eindtoestand na t stappen is.

Een Markov-strategie voor speler n voor het t-staps spel (notatie π_n^t)

is dan $\pi_n^t = \{\pi_n(k,\tau) \mid k=1,\ldots,N;\tau=0,\ldots,t-1\}$, waarbij $\pi_n(k,\tau)$ een gemengde aktie voor speler n in toestand k is, voor alle k en τ. De Markov-strategie π_n^t schrijft speler n voor om op tijdstip τ in toestand k gemengde aktie $\pi_n(k,\tau)$ te spelen. De totale verwachte uitbetaling in het t-stapsspel voor de strategieën π_1^t en π_2^t met k als begintoestand noteren we als $X^t(k,\pi_1^t,\pi_2^t)$. Laat $\bar{X}^t(\pi_1^t,\pi_2^t) = (X^t(1,\pi_1^t,\pi_2^t),\ldots,X^t(N,\pi_1^t,\pi_2^t))$.

Laat $\{\Gamma^t(k,\bar{R})\}$ het specifieke t-stapsspel met nabetalingsvektor $\bar{R}$, waarbij k de begintoestand is. Een matrix-spel, waarbij het $(i,j)^e$ element gelijk is aan g_{ij} noteren we ook wel als $\{g_{ij}\}$.

STELLING 1.2.1. *Ieder* t-*staps stochastisch spel van Shapley met nabetalingsvektor* $\bar{R}$ *bezit een waardevektor (notatie* $\bar{V}^*$*), die gevonden kan worden d.m.v. de volgende recurrente betrekking:*

$$\mathrm{Val}\ \{\Gamma^\tau(k,\bar{R})\} = \mathrm{Val}\ \{r(k,i,j) + \sum_{\ell=1}^{N} p(\ell|k,i,j)\cdot \mathrm{Val}\ \{\Gamma^{\tau-1}(\ell,\bar{R})\}\}$$
$$k = 1,2,\ldots,N;\ \tau = 1,2,\ldots,t,$$

waarbij $\mathrm{Val}\ \{\Gamma^0(k,\bar{R})\} = R(k)$.
V_k^* *is dan gelijk aan* $\mathrm{Val}\ \{\Gamma^t(k,\bar{R})\}$.
Beide spelers bezitten optimale Markov-strategieën. De Markov-strategie voor speler n, *die speler* n *voorschrijft om op tijdstip* τ *in toestand* k *een gemengde aktie* $\pi_n^*(k,\tau)$ *te kiezen, met* $\pi_n^*(k,\tau)$ *een optimale aktie voor speler* n *in het matrix-spel* $\{r(k,i,j) + \sum_{\ell=1}^N p(\ell|k,i,j).\ \mathrm{Val}\ \{\Gamma^{t-\tau-1}(\ell,\bar{R})\}\}$, *is optimaal voor speler* n.

BEWIJS. Het bewijs is gebaseerd op het principe van de volledige inductie.

a. Voor $t = 1$ hebben we alleen te maken met de volgende N matrix-spelen:

$$\{r(k,i,j) + \sum_{\ell=1}^{N} p(\ell|k,i,j)\cdot R\ (\ell)\},\ k = 1,\ldots,N.$$

Volgens stelling 1.1.1. hebben deze spelen een waarde en bezitten de spelers optimale strategieën, waarmee de stelling voor $t = 1$ bewezen is.

b. Veronderstel nu, dat de bewering juist is voor $\tau = 1,2,\ldots,t-1$.
We laten zien, dat de bewering dan ook juist is voor $\tau = t$.
Laat voor het (t-1)-stapsspel $\pi_n^{t-1^*} = \{\pi_n^*(k,\tau) \mid k = 1,\ldots,N;$

$\tau = 1,2,\ldots, t-2\}$ een optimale strategie voor speler n, n = 1,2.
Dan is $\bar{X}^{t-1}(\pi_1^{t-1^*}, \pi_2^{t-1^*})$ de waardevektor voor het (t-1)-stapsspel.
Beschouw nu de volgende N matrix spelen:

$$\{r(k,i,j) + \sum_{\ell=1}^{N} p(\ell|k,i,j)\cdot X^{t-1}(\ell, \pi_1^{t-1^*}, \pi_2^{t-1^*})\},\ K = 1,2,\ldots,N.$$

Laat $\bar{V}^* = (V_1^*,\ldots,V_N^*)$, waarbij V_k^* de waarde is van het k^e matrix spel en laat $\pi_n^*(k)$ hierbij een optimale aktie voor speler n, n = 1,2.
Definieer nu voor het t-stapsspel de Markov-strategie $\pi_n^{t^*}$ voor speler n als volgt: $\pi_n^{t^*}$ schrijft in de eerste stap $\pi_n^*(k)$ voor in toestand k, $\pi_n^{t^*}$ schrijft in de τ^e stap $\pi_n^*(k,\tau-2)$ voor in toestand k, $\tau = 2,\ldots,t$.
Laat voor een willekeurige gedragsstrategie π_2^G voor speler 2 het voorschrift in de eerste stap $\pi_2(k,1)$ in toestand k en definieer bij π_2^G de strategie $\pi_2^{G-1}(k,i,j)$ voor het (t-1)-stapsspel, waarbij $\pi_2^{G-1}(k,i,j)$ in de eerste stap in toestand ℓ voorschrijft wat π_2^G in de tweede stap in toestand ℓ voorschrijft bij historie $h_1 = (k,i,j,\ell)$, etc.
Dan geldt voor k = 1,2,...,N:

$$(1.1)\ \min_{\pi_2^G} X^t(k,\pi_1^{t^*},\pi_2^G) =$$

$$= \min_{\pi_2^G} \sum_{j=1}^{m_{k2}} \pi_2(k,1)(j) \sum_{i=1}^{m_{k1}} \pi_1^*(k)(i)\{r(k,i,j) +$$

$$+ \sum_{\ell=1}^{N} p(\ell|k,i,j)\ X^{t-1}(\ell,\pi_1^{t-1^*},\pi_2^{G-1}(k,i,j))\} \geq$$

$$\geq \min_{\pi_2^G} \sum_{j=1}^{m_{k2}} \pi_2(k,1)(j) \sum_{i=1}^{m_{k1}} \pi_1^*(k)(i)\{r(k,i,j) +$$

$$+ \sum_{\ell=1}^{N} p(\ell|k,i,j)\ X^{t-1}(\ell,\pi_1^{t-1^*},\pi_2^{t-1^*})\} = \mathrm{Val}\ \{r(k,i,j) +$$

$$+ \sum_{\ell=1}^{N} p(\ell|k,i,j)\cdot X^{t-1}(\ell,\pi_1^{t-1^*},\pi_2^{t-1^*})\} = V_k^*.$$

Dus

$$(1.2)\qquad \min_{\pi_2^G} X^t(k,\pi_1^{t^*},\pi_2^G) \geq V_k^* \qquad k = 1,\ldots,N.$$

Evenzo kunnen we laten zien, dat geldt:

$$(1.3 \qquad \max_{\pi_1^G} X^t(k,\pi_1^G,\pi_2^{t^*}) \le V_k, \; k = 1,\dots,N.$$

Kombineren van (1.2.) en (1.3.) geeft

$$(1.4 \qquad \bar{X}^t(\pi_1^G,\pi_2^{t^*}) \le \bar{X}^t(\pi_1^{t^*},\pi_2^{t^*}) = \bar{V}^* \le \bar{X}^t(\pi_1^{t^*},\pi_2^G), \quad \text{voor alle } \pi_1^G \text{ en } \pi_2^G.$$

Hieruit volgt $\sup_{\pi_1^G} \inf_{\pi_2^G} X^{-t}(\pi_1^G,\pi_2^G) = \inf_{\pi_2^G} \sup_{\pi_1^G} \bar{X}(\pi_1^G,\pi_2^G) = \bar{V}^*$, zodat het spel een waardevektor bezit, die gelijk is aan $\bar{V}^* = \bar{X}^t(\pi_1^{t^*},\pi_2^{t^*})$.
Uit (1.4) blijkt, dat $\pi_1^{t^*}$ en $\pi_2^{t^*}$ optimale strategieën zijn voor respektievelijk speler 1 en speler 2.

Uit (1.1 blijkt de recurrente betrekking voor $\tau = t$. Uit de constructiewijze van $\pi_1^{t^*}$ en $\pi_2^{t^*}$ en de aanname, dat $\pi_1^{t-1^*}$ en $\pi_2^{t-2^*}$ Markov-strategieën zijn, blijkt dat ook $\pi_1^{t^*}$ en $\pi_2^{t^*}$ Markov-strategieën zijn.
Hiermede is de inductiestap voltooid, waarmee de stelling bewezen is. □

We gaan nu het stochastisch spel van Shapley analyseren.
De notaties voor de diverse soorten strategieën zijn gelijk aan die ingevoerd in sektie 1.1.
Voor een paar stationaire strategieën π_1 en π_2 kunnen we voor de totale verwachte uitbetaling de volgende betrekking opstellen:

$$(1.5\,) \qquad X(k,\pi_1,\pi_2) = r\,(k,\pi_1,\pi_2) + \sum_{\ell=1}^{N} p(\ell|k,\pi_1,\pi_2)\cdot X(\ell,\pi_1,\pi_2) \qquad k=1,\dots,N.$$

Hierbij is

$$r(k,\pi_1,\pi_2) = \sum_{i=1}^{m_{k1}} \sum_{j=1}^{m_{k2}} \pi_1(k,i)\cdot\pi_2(k,j)\cdot r(k,i,j)$$

en

$$p(\ell|k,\pi_1,\pi_2) = \sum_{i=1}^{m_{k1}} \sum_{j=1}^{m_{k2}} \pi_1(k,i)\cdot\pi_2(k,j)\cdot p(\ell|k,i,j).$$

<u>DEFINITIE</u>. De $m_{k1} \times m_{k2}$ matrix spelen $G_k(\bar{v})$, waarbij het $(i.j)^e$ element

gelijk is aan $r(k,i,j) + \sum_{\ell=1}^{N} p(\ell|k,i,j)\, v_\ell$ worden de dummy spelen bij de N-vektor $\bar{v}$ genoemd, $k = 1,\dots,N$.

STELLING 1.2.2. *Ieder stochastisch spel van Shapley bezit voor het totale opbrengstenmodel een waardevektor, die gevonden kan worden als de unieke oplossing* $\bar{v}^*$ *van het stelsel* $v_k = \mathrm{Val}\ G_k(\bar{v})$, $k = 1,\dots,N$.

BEWIJS. Definieer de afbeelding $T : \mathbb{R}^N \to \mathbb{R}^N$ als $T\bar{\alpha} = \bar{\beta}$ met $\beta_k = \mathrm{Val}\ G_k(\bar{\alpha})$. Definieer op $\mathbb{R}^N$ de norm $\|\bar{\alpha}\| = \max_k |\alpha_k|$, dan is de $\mathbb{R}^N$ een volledige metrische ruimte onder deze norm.
We bewijzen nu dat T een contractieafbeelding is onder de norm $\|\cdot\|$. Voor alle $\bar{\alpha}, \bar{\beta} \in \mathbb{R}^N$ vinden we met behulp van lemma 1.1.3:

$$\begin{aligned}
\|T\bar{\beta} - T\bar{\alpha}\| &= \max_k \left| \mathrm{Val}\ G_k(\bar{\beta}) - \mathrm{Val}\ G_k(\bar{\alpha}) \right| \\
&\le \max_{k,i,j} \left| \sum_{\ell=1}^{N} P(\ell|k,i,j)(\beta_\ell - \alpha_\ell) \right| \\
&\le \max_{k,i,j} \left| \sum_{\ell=1}^{N} p(\ell|k,i,j) \right| \max_\ell \left| \beta_\ell - \alpha_\ell \right| \\
&= (1 - q_0)\ \|\bar{\beta} - \bar{\alpha}\|,
\end{aligned}$$

waarmee bewezen is dat T een contractieafbeelding is met contractiefactor $\le 1 - q_0$.
Uit de vaste punten stelling volgt nu dat T een uniek dekpunt $\bar{v}^*$ heeft, waarbij voor elke $\bar{\alpha} \in \mathbb{R}^N$ geldt $\lim_{t\to\infty} \|T^t\bar{\alpha} - \bar{v}^*\| = 0$, waaruit volgt $\lim_{t\to\infty} T^t\bar{\alpha} = \bar{v}^*$, hierbij is $T^t(\bar{\alpha}) = T(T^{t-1}(\bar{\alpha}))$, $t = 2,3,\dots$. Beschouw nu het t-staps spel met nabetalingsvektor $\bar{0}$ (N-vektor met iedere komponent 0).
Uit de definitie van dummy spelen, de definitie van T en stelling 1.2.1 blijkt dan dat $T^t(\bar{0})$ de waardevektor van het t-staps spel met nabetalingsvektor $\bar{0}$ is.
Laat $\bar{1}$ de N-vektor zijn met iedere komponent gelijk aan 1. Als nu in het stochastische spel speler 1 een strategie speelt, die wat de eerste t stappen betreft overeenstemt met een optimale strategie voor het t-stapsspel, dan krijgt speler 1 een totale verwachte uit betaling van tenminste $T^t(\bar{0}) - \varepsilon_t.\bar{1}$, waarbij $\varepsilon_t = (1 - q_0)^t \frac{M}{q_0}$, tegen iedere strategie van speler 2. (ε_t is een

begrenzing voor de totale verwachte uitbetaling na tijdstip t). Daar $\varepsilon_t \to 0$ en $T^t(\bar{0}) \to \bar{v}^*$ als $t \to \infty$ zien we dat er voor iedere $\varepsilon > 0$ een π_1^ε bestaat, z.d.d.

$$\bar{X}(\pi_1^\varepsilon,\pi_2^G) \geq \bar{v}^* - \varepsilon.\bar{1} \qquad \text{voor alle} \qquad \pi_2^G \in \Pi_2^G,$$

waaruit volgt $\inf_{\pi_2^G} \bar{X}(\pi_1^\varepsilon,\pi_2^G) \geq \bar{v}^* - \varepsilon.\bar{1}$, zodat

(1.6) $$\sup_{\pi_1^G} \inf_{\pi_2^G} \bar{X}(\pi_1^G,\pi_2^G) \geq \bar{v}^*.$$

Evenzo kunnen we laten zien

(1.7) $$\inf_{\pi_2^G} \sup_{\pi_1^G} \bar{X}(\pi_1^G,\pi_2^G) \leq \bar{v}^*.$$

Eenvoudig is in te zien dat voor iedere vektorfunctie $\bar{V}\colon \pi_1^G \times \pi_2^G \to \mathbb{R}^N$ geldt

(1.8) $$\sup_{\pi_1^G} \inf_{\pi_2^G} \bar{V}(\pi_1^G,\pi_2^G) \leq \inf_{\pi_2^G} \sup_{\pi_1^G} \bar{V}(\pi_1^G,\pi_2^G).$$

Kombineren van (1.6), (1.7) en (1.8) geeft

$$\sup_{\pi_1^G} \inf_{\pi_2^G} \bar{X}(\pi_1^G,\pi_2^G) = \inf_{\pi_2^G} \sup_{\pi_1^G} \bar{X}(\pi_1^G,\pi_2^G) = \bar{v}^*,$$

waarmede de stelling bewezen is. □

STELLING 1.2.3. *In een spel van Shapley hebben beide spelers optimale stationaire strategieën, waarbij het voorschrift voor speler* n *in toestand* k *overeenstemt met een optimale aktie voor speler* n *in het dummy matrixspel* $G_k(\bar{v}^*)$ *met* $\bar{v}^*$ *de waarde van het spel.*

BEWIJS. Beschouw het t-staps spel met nabetalingsvektor $\bar{v}^*$. Aangezien $v_k^* = \text{Val}\ G_k(\bar{v}^*)$, $k = 1,\ldots,N$ zien we met induktie dat $\text{Val}\{\Gamma^\tau(k,\bar{v}^*)\} = v_k^*$, $\tau = 1,2,\ldots,t$; $k = 1,\ldots,N$.

Een optimale Markov-strategie π_n^{t*} voor speler n voor dit t-staps spel overeenkomstig stelling 1.2.1 is dan: $\pi_n^{t*} = \{\pi_n^*(h,\tau) \mid k = 1,\ldots,N;\ \tau = 0,1,\ldots,t-1\}$ waarbij $\pi_n^*(k,\tau) = \pi_n^*(k)$, $\tau = 0,\ldots,t-1$ met $\pi_n^*(k)$ een optimale aktie voor speler n in $G_k(\bar{v}^*)$.
Speelt speler 1 nu de stationaire strategie $\pi_1^* = (\pi_1^*(1),\ldots,\pi_1^*(N))$ in het stochastische spel, dan geldt voor iedere t dat de totale verwachte uitbetaling aan speler 1 na t stappen minstens $\bar{v}^* - \varepsilon_t \cdot \bar{1}$ is tegen iedere strategie van speler 2, met $\varepsilon_t \le (1-p_0)^t \max_\ell |v_\ell^*|$. Daar de totale verwachte uitbetaling na tijdstip t naar 0 gaat als $t \to \infty$ en daar $\lim_{t\to\infty} \varepsilon_t = 0$ volgt nu direkt dat $\bar{X}(\pi_1^*,\pi_2^G) \ge \bar{v}^*$ voor alle $\pi_2^G \in \Pi_2^G$.
Evenzo vinden wij $\bar{X}(\pi_1^G,\pi_2^*) \le \bar{v}^*$, voor alle $\pi_1^G \in \Pi_1^G$, waarmee de stelling bewezen is. □

<u>DEFINITIE</u>. Voor $\varepsilon > 0$ is een ε-optimale strategie voor speler n een strategie die, tegen iedere strategie van de andere speler, hem een totale verwachte uitbetaling geeft die hoogstens ε slechter is voor hem dan de waarde van het spel.

<u>LEMMA 1.2.4</u>. *Een stationaire strategie* $\pi_n^\varepsilon = (\pi_n^\varepsilon(1),\ldots,\pi_n^\varepsilon(N))$ *voor speler* n *waarbij* $\pi_n^\varepsilon(k)$ *overeenstemt met een* ε*-optimale gemengde aktie voor speler* n *in matrixspel* $G_k(\bar{v}^*)$ *is* ε/q_0 *optimaal voor speler* n *in het stochastische spel.*

<u>BEWIJS</u>. We bewijzen het lemma voor speler 1. Laat π_1^ε als in het lemma gesteld. Voor π_1^ε en een stationaire strategie π_2 voor speler 2 geldt voor de totale verwachte opbrengst volgens (1.5):

$$(1.9) \qquad \bar{X}(\pi_1^\varepsilon,\pi_2) = \bar{r}(\pi_1^\varepsilon,\pi_2) + P_{\pi_1^\varepsilon\pi_2} \cdot \bar{X}(\pi_1^\varepsilon,\pi_2).$$

Daar $\pi_1^\varepsilon(k)$ ε-optimaal is voor speler 1 in het matrixspel $G_k(\bar{v})$ geldt:

$$(1.10) \qquad \bar{r}(\pi_1^\varepsilon,\pi_2) + P_{\pi_1^\varepsilon\pi_2} \cdot \bar{v}^* \ge \bar{v}^* - \varepsilon\cdot\bar{1}.$$

Kombineren van (1.9) en (1.10) geeft:

$$(1.11) \qquad \bar{X}(\pi_1^\varepsilon,\pi_2) - \bar{v}^* \ge P_{\pi_1^\varepsilon\pi_2} \cdot (\bar{X}(\pi_1^\varepsilon,\pi_2) - \bar{v}^*) - \varepsilon\bar{1}.$$

Als we nu het linkerlid van (1.11) herhaald invullen in het rechterlid, zien we dat voor ieder t geldt:

$$\bar{X}(\pi_1^\varepsilon,\pi_2) - \bar{v}^* \geq P^t_{\pi_1^\varepsilon\pi_2}\,(\bar{X}(\pi_1^\varepsilon,\pi_2) - \bar{v}^*) - \varepsilon \cdot \sum_{\tau=0}^{t-1} P^\tau_{\pi_1^\varepsilon\pi_2} \cdot \bar{1}.$$

Daar $\lim_{t\to\infty} P^t_{\pi_1^\varepsilon\pi_2}$ gelijk is aan een matrix bestaande uit nullen volgt hieruit:

$$\bar{X}(\pi_1^\varepsilon,\pi_2) \geq \bar{v}^* - \varepsilon \sum_{\tau=0}^{\infty} P^\tau_{\pi_1^\varepsilon\pi_2} \cdot \bar{1} \geq \bar{v}^* - \varepsilon \sum_{\tau=0}^{\infty} (1-q_0)^\tau \cdot \bar{1} = \bar{v}^* - \frac{\varepsilon}{q_0} \cdot \bar{1}.$$

Daar π_2 een willekeurige stationaire strategie voor speler 2 is, is hiermee het lemma voor de klasse der stationaire strategiën bewezen. Als speler 1 een stationaire strategie speelt, bekend aan speler 2, dan ziet speler 2 zich geconfronteerd met een Markov-beslissingsprobleem. In de theorie der Markov-beslissingsproblemen is bekend dat binnen de klasse der stationaire strategieën een optimale strategie gevonden kan worden (bijvoorbeeld BLACKWELL [3]). Door dit resultaat te gebruiken verkrijgen we:

$$\inf_{\pi_2^G\in\Pi_2^G} \bar{X}(\pi_1^\varepsilon,\pi_2^G) = \inf_{\pi_2\in\Pi_2} \bar{X}(\pi_1^\varepsilon,\pi_2) \geq \bar{v}^* - \frac{\varepsilon}{q_0} \cdot \bar{1},$$

waarmee de stelling bewezen is. □

In het bewijs van stelling 1.2.2 hebben we de afbeelding T: $\mathbb{R}^N \to \mathbb{R}^N$ gedefinieerd als $T\bar{\alpha} = \bar{\beta}$ met $\beta_k = \mathrm{Val}\ G_k(\bar{\alpha})$, waarbij $G_k(\bar{\alpha})$ het k^e dummy spel bij $\bar{\alpha}$ is.

STELLING 1.2.5. *De methode der successieve approximatie door middel van de afbeelding* T *kan voor iedere* $\varepsilon > 0$ *zowel de waarde in een eindig aantal iteraties tot op* ε *benaderen, als in een eindig aantal iteraties* ε*-optimale stationaire strategieën voor de beide spelers opleveren.*

BEWIJS. Laat $\varepsilon > 0$. Laat $\bar{v}^*$ de waarde van het spel zijn. In het bewijs van stelling 1.2.2 is bewezen $T\bar{v}^* = \bar{v}^*$ en $\|T\bar{\alpha}-T\bar{\beta}\| \leq (1-q^0)\|\bar{\alpha}-\bar{\beta}\|$, $\forall \bar{\alpha},\bar{\beta} \in \mathbb{R}^N$. Hieruit volgt $\|T^t\bar{\alpha}-T^t\bar{\beta}\| \leq (1-q_0)^t\|\bar{\alpha}-\bar{\beta}\|$. Merk op dat $\|\bar{v}^*\| \leq M/q_0$.
Neem $\bar{\alpha} = \bar{0}$ en $\bar{\beta} = \bar{v}^*$, dan geldt:

$$\|T^t\bar{0}-T^t\bar{v}^*\| = \|T^t\bar{0}-\bar{v}^*\| \le (1-q_0)^t\|\bar{v}^*\| \le (1-q_0)^t \cdot \frac{M}{q_0}.$$

Door nu t_ε zodanig te kiezen dat $(1-q_0)^{t_\varepsilon} \cdot M/q_0 \le \varepsilon$ geldt, dus

$$\max_k |T_k^{t_\varepsilon}(\bar{0}) - v_k^t| \le \varepsilon,$$

waarmee het eerste deel van de stelling bewezen is.
Laat nu $\pi_n^\varepsilon = (\pi_n^\varepsilon(1),\ldots,\pi_n^\varepsilon(N))$ de stationaire strategie voor speler n, waarbij $\pi_n^\varepsilon(k)$ overeenstemt met een optimale aktie voor speler n in het dummy spel $G_k(T^{t_\varepsilon}(\bar{0}))$, n = 1,2. Dan geldt

$$(1.12)\qquad \bar{r}(\pi_1^\varepsilon,\pi_2) + P_{\pi_1^\varepsilon\pi_2}\cdot T^{t_\varepsilon}(\bar{0}) \ge T^{t_\varepsilon+1}(\bar{0}) \ge \bar{v}^* - \varepsilon\cdot\bar{1}$$

voor iedere stationaire strategie voor speler 2.
Daar $T^{t_\varepsilon}(\bar{0}) \le \bar{v}^* + \varepsilon$ volgt uit (1.12):

$$(1.13)\qquad \bar{r}(\pi_1^\varepsilon,\pi_2) + P_{\pi_1^\varepsilon\pi_2}\bar{v}^* \ge \bar{v}^* - \varepsilon(I + P_{\pi_1^\varepsilon\pi_2})\cdot\bar{1}$$

$$\ge \bar{v}^* - \varepsilon\cdot(2-q_0)\cdot\bar{1}.$$

Hierbij is I de eenheidsmatrix.
Uit (1.13) blijkt nu dat π_1^ε zodanig dat $\pi_1^\varepsilon(k)$ overeenstemt met een $\varepsilon(2-q_0)$-optimale aktie voor speler 1 in $G_k(v^*)$. Uit lemma 1.2.4 volgt dan dat π_1^ε $\frac{\varepsilon(2-q_0)}{q_0}$-optimaal is voor speler 1 in het stochastische spel. Evenzo kunnen we aantonen dat π_2^ε $\frac{\varepsilon(2-q_0)}{q_0}$-optimaal is voor speler 2. Hiermee is de stelling bewezen. □

Enkele successieve approximatiemethoden staan uitgewerkt in VAN DER WAL [21] en [22], waarin verbeterde grenzen voor de benadering worden gegeven.

<u>OPMERKINGEN.</u>

1. Het stochastisch spel van Shapley met rationale coëfficiënten r(k,i,j) en $p(\ell|k,i,j)$ hoeft niet noodzakelijkerwijs een rationale waarde te

hebben. Dit in tegenstelling tot de Markov-beslissingsproblemen met een eindig aantal toestanden en eindig aantal akties in iedere toestand.

VOORBEELD. 1 toestand, beide spelers hebben twee zuivere akties. Het dummy spel $G_1(v)$ ziet er als volgt uit

$$\begin{pmatrix} 4+0\cdot v & 1+\frac{1}{2}v \\ 2+0\cdot v & 1+\frac{3}{4}v \end{pmatrix}.$$

De unieke oplossing van de waardevergelijking $v = \mathrm{Val}\begin{pmatrix} 4 & 1+\frac{1}{2}v \\ 2 & 1+\frac{3}{4}v \end{pmatrix}$ wordt gegeven door $v^* = \sqrt{8}$.

2. Een niet-stoppend stochastisch spel, waarbij de opbrengsten worden verdisconteerd met een verdisconteringsfactor β, $0 \le \beta < 1$, is op te vatten als een speciaal geval van het spel van Shapley.
Beschouw bij een niet-stoppend spel namelijk het spel waarin de overgangskansen gelijk zijn aan $\beta\cdot p(\ell|k,i,j)$, alle k,ℓ,i,j, Hiermee hebben we een stoppend spel verkregen, min $1 - \sum_{\ell=1}^{N} p(\ell|k,i,j) = 1-\beta$. Het is gemakkelijk in te zien, dat voor dit nieuwe spel onder ieder paar strategieën voor de beide spelers de totale verwachte opbrengst precies gelijk is aan de totale verwachte verdisconteerde opbrengst voor deze beide strategieën in het oorspronkelijke spel. Alle resultaten ontwikkeld in deze sektie zijn dus ook van toepassing op niet-stoppende spelen onder het verdisconteerde criterium.

3. Een stochastisch spel van Shapley met perfekte informatie is een spel, waarbij in toestand k óf $m_{k1} = 1$ óf $m_{k2} = 1$, $k = 1,2,\dots,N$. Uit stelling 1.2.3 blijkt dan dat voor een spel met perfekte informatie beide spelers optimale zuivere stationaire strategieën bezitten.

4. Als $m_{k2} = 1$, $k = 1,2,\dots,N$, dus de tweede speler heeft in iedere toestand maar 1 aktie ter beschikking, waardoor hij in wezen geëlimineerd wordt, dan hebben we te maken met een Markov-beslissingsprobleem voor speler 1. Dynamische programmeringstechnieken geven ons dan de oplossing van het probleem (zie bijvoorbeeld HOWARD [7]). In dit geval bezit speler 1 optimale zuivere stationaire strategieën, hetgeen direkt uit opmerking 3 volgt, daar we hier te maken hebben met een speciaal geval van een spel met perfekte informatie.

1.3. Twee-persoons niet-nulsom stochastische spelen

In deze sektie geven we enkele resultaten betreffende twee-persoons niet-nulsom spelen. Even als in het spel van Shapley veronderstellen we dat we te maken hebben met een eindig, stoppend stochastisch spel. Het enige verschil met het model uit de vorige sektie is nu dat iedere speler zijn eigen uitbetalingsfunctie heeft.
Het zoeken is niet langer naar optimale strategieën, maar naar evenwichtspunten (voor een definitie, zie sektie 1.1).
We veronderstellen het begrip bimatrix-spel bekend. De notaties, die gebruikt zullen worden zijn dezelfde als in de vorige sektie met als aanvulling de uitbetalingsfuncties $r_1(k,i,j)$ en $r_2(k,i,j)$. We zullen ons in deze sektie beperken tot de klasse der stationaire strategieën. Het volgende lemma, dat we niet zullen bewijzen, geeft aan dat dit in bepaalde mate geen restrictie geeft.

LEMMA 1.3.1. *Een evenwichtspunt onder het totale opbrengstencriterium voor een eindig stoppend tweepersoons niet-nulsom stochastisch spel binnen de klasse der stationaire strategieën is ook evenwichtspunt binnen de klasse van alle gedragsstrategieën.*

Het bewijs van dit lemma is gebaseerd op het volgende: Voor vaste stationaire strategie van een speler, die bekend is aan de ander, kijkt deze tegen een Markov-beslissingsprobleem aan en voor Markov-beslissingsproblemen is bekend dat optimale strategieën gevonden kunnen worden binnen de klasse der stationaire strategieën.

Een stationaire strategie wordt in het vervolg kortweg aangeduid als strategie.

Laat Π_n de verzameling van stationaire strategieën voor speler n zijn. Een element $\pi_n = (\pi_n(1),\ldots,\pi_n(N)) \in \Pi_n$ kan worden opgevat als een element uit de $\mathbb{R}^{I_n}$ met $I_n = \sum_{k=1}^{N} m_{kn}$.
Aldus is na te gaan dat Π_n opgevat kan worden als een kompakte convexe deelverzameling van de $\mathbb{R}^{I_n}$.
We beschouwen evenals in de vorige sektie het totale opbrengstenmodel.

STELLING 1.3.2. *Ieder eindig stoppend twee-persoons niet-nulsom stochastisch spel bezit een evenwichtspunt binnen de klasse der stationaire strategieën.*

BEWIJS. We geven geen strikt bewijs daar dit te ver zou voeren voor de bedoeling van dit rapport.
In de eerste plaats merken we op dat $\bar{X}_n(\pi_1,\pi_2)$ een continue vektorfunctie is op $\Pi_1 \times \Pi_2$. Dit is gebaseerd op het feit, dat de staartopbrengst uniform naar 0 gaat voor ieder paar strategieën, hetgeen een gevolg is van het stoppend zijn van het spel.
Definieer nu de volgende punt naar verzameling afbeelding T.
T voegt aan ieder paar $(\pi_1^0,\pi_2^0) \in \Pi_1\times\Pi_2$ een deelverzameling $T(\pi_1^0,\pi_2^0) \subset \subset \Pi_1\times\Pi_2$ toe en wel zodanig dat $(\pi_1,\pi_2) \in T(\pi_1^0,\pi_2^0)$, dan en slechts dan als $\bar{X}_1(\pi_1,\pi_2^0) = \max_{\pi_1\in\Pi_1} \bar{X}_1(\pi_1,\pi_2^0)$ en $\bar{X}_2(\pi_1^0,\pi_2) = \max_{\pi_2\in\Pi_2} \bar{X}_2(\pi_1^0,\pi_2)$.
We kunnen laten zien dat $T(\pi_1^0,\pi_2^0)$ niet leeg is voor alle $(\pi_1^0,\pi_2^0) \in \Pi_1\times\Pi_2$, verder is $T(\pi_1^0,\pi_2^0)$ een kompakte verzameling en met behulp van de continuïteit van $\bar{X}_n(\pi_1,\pi_2)$ volgt dat de afbeelding T boven-semi-continu is, d.w.z. hebben wij een rij $\{(\pi_1,\pi_2)_n\}$ in $\Pi_1\times\Pi_2$ die convergeert naar $(\pi_1,\pi_2)_0$ en als $(\pi_1^1,\pi_2^1)_n$ een element van $T(\pi_1,\pi_2)_n$ is, dan is ieder limietpunt van de rij $\{(\pi_1^1,\pi_2^1)_n\}$ een element van $T(\pi_1,\pi_2)_0$.
We kunnen nu de stelling van KAKUTANI [8] toepassen, die zegt dat iedere boven-semi-continue punt naar kompakte verzameling afbeelding van een kompakt convexe metrische verzameling in zichzelf minstens één dekpunt bezit.
We concluderen dat de afbeelding T minstens één dekpunt bezit. Laat (π_1^*,π_2^*) zo'n dekpunt zijn. Er geldt dan

$$\bar{X}_1(\pi_1,\pi_2^*) \le \max_{\pi_1\in\Pi_1} \bar{X}_1(\pi_1,\pi_2^*) = \bar{X}_1(\pi_1^*,\pi_2^*)$$

en

$$\bar{X}_2(\pi_1^*,\pi_2) \le \max_{\pi_2\in\Pi_2} \bar{X}_2(\pi_1^*,\pi_2) = \bar{X}_2(\pi_1^*,\pi_2^*),$$

waaruit blijkt, dat (π_1^*,π_2^*) een evenwichtspunt vormt. □

Het resultaat in stelling 1.3.2 is het eerst bewezen door ROGERS [12].

Nadien hebben o.a. FEDERGRÜN [6] en VRIEZE [19] soortgelijke resultaten bewezen voor stochastische spelen met uitgebreidere spelersverzamelingen, toestandsruimten en aktieruimten.

We vermelden als afsluiting van dit hoofdstuk nog een lemma, dat behulpzaam kan zijn bij het construeren van evenwichtspunten voor niet-nulsom stochastische spelen.

DEFINITIE. Voor een twee-persoons niet-nulsom stochastisch spel worden voor $\bar{v}_1, \bar{v}_2 \in \mathbb{R}^N$ de volgende N bimatrix-spelen de *dummy bimatrix-spelen* bij $\bar{v}_1$ en $\bar{v}_2$ genoemd:

$$G_k(\bar{v}_1,\bar{v}_2) = (A_k(\bar{v}_1), B_k(\bar{v}_2)), \qquad k = 1,\dots,N,$$

waarbij

$$a_k(i,j) = r_1(k,i,j) + \sum_{\ell=1}^{N} p(\ell|k,i,j)\cdot v_1(\ell)$$

en

$$b_k(i,j) = r_2(k,i,j) + \sum_{\ell=1}^{N} p(\ell|k,i,j)\cdot v_2(\ell).$$

LEMMA 1.3.3. *Een paar stationaire strategieën* $\pi_1^* = (\pi_1^*(1),\dots,\pi_1^*(N))$ *en* $\pi_2^* = (\pi_2^*(1),\dots,\pi_2^*(N))$ *vormt dan en slechts dan een evenwichtspunt voor een niet-nulsom stochastisch spel, als het paar* $(\pi_1^*(k),\pi_2^*(k))$ *overeenstemt met een evenwichtspunt voor het dummy-bimatrix-spel*

$$G_k(\bar{X}_1(\pi_1^*,\pi_2^*),\bar{X}_2(\pi_1^*,\pi_2^*)), \qquad k = 1,\dots,N.$$

Het bewijs van dit lemma kan worden gevonden in VRIEZE [19]. Uit het lemma blijkt, dat algoritmen die evenwichtspunten voor bimatrixspelen opleveren van nut kunnen zijn voor het vinden van evenwichtspunten voor niet-nulsom stochastische spelen.

2. TOEPASSING VAN DE THEORIE DER STOCHASTISCHE SPELEN OP ZOEKMODELLEN

2.1 INLEIDING

We formuleren eerst twee problemen om duidelijk te maken wat met zoekproblemen bedoeld wordt.

a. Een surveilleer probleem.

Het politieapparaat ziet zich voor het volgende probleem gesteld. Het is hun bekend geworden, dat gedurende een bepaalde periode een inbrekersbende zich wil verrijken d.m.v. het kraken van banken. Hiervoor komen N banken in aanmerking. Iedere nacht wordt hoogstens één kraak gezet. Er zijn te weinig politiemensen om alle banken te bewaken en de kans van slagen van een kraak op een bewaakte bank hangt af van het aantal bewakende politiemensen.

De politie vraagt zich af welke banken en met hoeveel mensen ze moet bewaken. Evenzo vragen de inbrekers zich af welke banken ze moeten kraken en wanneer dit moet gebeuren.

b. Een medisch probleem.

Een bepaalde ziekte kan N oorzaken hebben. Een medisch onderzoeker heeft een bepaald budget B (in de vorm van tijd,medewerkers,technisch faciliteiten,geld) ter beschikking om de oorzaak op te sporen. Uit statistisch materiaal kan hij opmaken, dat oorzaak i met kans x_i optreedt. Bij oorzaak i zijn er m_i testen, waarvan de j^e test met kans $p_i(j)$ een positief resultaat geeft als i de oorzaak is en een negatief resultaat als i niet de oorzaak is.De kosten van dit j^e experiment voor testen op i zijn $c_i(j)$.

De medisch onderzoeker vraagt zich af hoe hij zijn budget moet besteden, opdat zijn kans om de oorzaak te vinden zo groot mogelijk is.

2.2 HET ALGEMENE ZOEKMODEL.

We gaan nu het algemene zoekmodel beschrijven. Deze valt uiteen in twee hoofdmodellen, die beiden opgesplitst kunnen worden in twee typen.

Algemeen. Gegeven zijn N doosjes genummerd 1 t/m N. Speler 1 mag zich één of meerdere keren in een doosje verstoppen en speler 2 moet uitzoeken in welk doosje speler 1 zich bevindt. Dit kan hij doen door achtereenvolgens in de doosjes te kijken. Als speler 2 in het goede doosje kijkt, is er een positieve kans, dat hij speler 1 niet ziet, zodat het heel wel mogelijk is, dat hij meerdere keren in hetzelfde doosje moet zoeken. Een periode in het probleem komt overeen met het inspekteren van speler 2 van een doosje.

Model 1.	Speler 1 verstopt zich 1 keer met bekende kansverdeling voor speler 2.	—a. eindige horizon bijv. eindig budget —b. oneindige horizon bijv. spel stopt alleen na vinden
Model 2.	Speler 1 verstopt zich na iedere periode opnieuw met onbekende kansverdeling voor speler 2.	—a. eindige horizon bijv.met kans 1 is speler 1 na een eindig aantal perioden ontsnapt of zit in de val —b. oneindige horizon bijv. alleen stoppen na vinden

Als speler 2 in doos i kijkt, dan maakt hij kosten c_i (eventueel te kiezen uit een eindig aantal als hij op verschillende manieren doos i kan onderzoeken), dan heeft hij een vindkans p_i als speler 1 zich in doos i bevindt (afhankelijk c_i indien c_i variabel) en heeft hij bij vinden een opbrengst r_i.

In alle gevallen veronderstellen we $p_i < 1$

In de modellen met een oneindige horizon is gespecificeerd een verdisconteringsfactor β , $0 \leq \beta < 1$.

Bovenstaande modellen kunnen we opvatten als een nulsom model; speler 1 probeert zich zo slim mogelijk te verweren tegen speler 2 zijn zoekakties; dus wat speler 2 aan moeite moet doen om speler 1 te vinden kan worden gezien als opbrengst voor speler 1.

De doelstellingen voor speler 2 kunnen velerlei zijn:

1^{o} max E {vindkans}

2^o min E {kosten tot vinden}

3^o min E {vindtijd}

4^o {doelstelling i^o onder de conditie doelstelling j^o begrensd},
$i,j = 1,2,3 \; ; i \neq j$.

De meest voorkomende bij 4^o is:

max E {vindkans} onder de conditie dat de kosten $\leq B$ zijn.

Evenzo kan speler 1 verschillende doelstellingen hebben.

2.3 HET ZOEKMODEL ALS STOCHASTISCH SPEL.

In deze sektie gaan we bovenstaande modellen als stochastische spelen formuleren.

Model 1^a. Laat de verstopkansverdeling voor speler 1 zijn $\bar{x}=(x_1,\ldots,x_N)$, d.w.z. speler 1 verstopt zich met kans x_i in doos $i, i=1,\ldots,N$.
We veronderstellen dat speler 2 $\bar{x}$ kent.
Laat $Q = \{\bar{q} \mid \bar{q}=(q_1,\ldots,q_N),\ q_i \geq 0,\ i=1,\ldots,N;\ \sum_{i=1}^{N} q_i = 1\}$
We veronderstellen, dat het spel hoogstens t perioden kan duren.
Als toestandsruimte in het stochastische spel nemen we $S = Q \cup \{0,1\}$.
Hierbij duidt 1 op de toestand dat speler 1 gevonden is, d.w.z. zo gauw speler 1 gevonden is springt het systeem naar toestand 1 en stopt het spel; 0 is de eindtoestand als speler 1 na t perioden nog niet gevonden is.
Q komt overeen met de verzameling van conditionele verstopkansverdelingen, die speler 2 kan ontmoeten. In periode 1 is deze kansverdeling gelijk aan de gegeven $\bar{x} = (x_1,\ldots,x_N)$. Als nu in periode 1 speler 2 zoekt in doos i met inspanning $c_i(j)$ en hij vindt speler 1 niet, dan geldt:

Pr {speler 1 bevindt zich in doos k | speler 2 heeft zonder succes in periode 1 in doos i gekeken} =

=Pr {speler 2 heeft zonder succes in periode 1 in doos i gekeken | speler 1 bevindt zich in doos k} .

$$\cdot \frac{\text{Pr \{speler 1 bevindt zich in doos k\}}}{\text{Pr \{speler 2 heeft zonder succes in periode 1 in doos i gekeken\}}} =$$

$$= 1 \cdot \frac{x_k}{1-p_i(c_i(j)).x_i} \qquad \text{voor } k \neq i \;,$$

$$= (1-p_i(c_i(j))) \cdot \frac{x_i}{1-p_i(c_i(j)) \cdot x_i} \qquad \text{voor } k=i.$$

Aldus doorgaande kunnen we voor iedere periode τ en rij van mislukte inspekties in de doosjes $i_1, i_2, \ldots, i_{\tau-1}$ een conditionele verstopkansverdeling van speler 1 voor periode τ berekenen.
Als speler 2 in toestand $\bar{q} \in Q$ aktie $c_i(j)$ kiest, waarmee we bedoelen, doos i onderzoeken met kosten $c_i(j)$, dan zijn zijn direkte verwachte kosten:

$$c_i(j) - p_i(c_i(j)) \cdot q_i \cdot r_i \quad .$$

De overgangskansen bij zuivere aktie $c_i(j)$ in toestand $\bar{q}$ zijn:

$$p(1 \mid \bar{q}, c_i(j)) = p_i(c_i(j)) \cdot q_i$$

$$p(\bar{q}' \mid \bar{q}, c_i(j)) = 1 - p_i(c_i(j)) \cdot q_i \quad \text{als } \bar{q}' = \bar{q}^* \text{met}$$

$$\bar{q}_i^* = \frac{(1-p_i(c_i(j))) \cdot q_i}{1-p_i(c_i(j)) \cdot q_i} \quad \text{en} \quad q_k^* = \frac{q_k}{1-p_i(c_i(j)) \cdot q_i} \,, k \neq i$$

$$p(\bar{q}' \mid \bar{q}, c_i(j)) = 0 \quad \text{als } \bar{q}' \neq \bar{q}^* \,.$$

Deze overgangskansen gelden voor $\tau < t$. Voor $\tau = t$ springt het spel met kans $p_i(c_i(j)) \cdot q_i$ naar toestand 1 en met kans $1 - p_i(c_i(j)) \cdot q_i$ naar toestand 0. Hiermee is het stochastisch spel geformuleerd.
Terwille van de eenvoud nemen we in het vervolg aan dat speler 2 ieder doosje maar op 1 manier kan onderzoeken. Voor iedere i is er dan maar één mogelijke c_i en $p_i(c_i)$ noteren we in het vervolg als p_i.

Laat $\Pi_2 = \{\bar{\pi} \mid \bar{\pi} = (\pi(1), \ldots, \pi(N)),\ \pi(i) \geq 0,\ i=1, \ldots, N;\ \sum_{i=1}^{N} \pi(i) = 1\}$ zijn.
Π_2 is de verzameling van gemengde akties voor speler 2 in een toestand $\bar{q} \in Q$.
Daar speler 1 een voorgeschreven strategie $\bar{x}$ speelt, hebben we eigenlijk te maken met een Markov-beslissingsprobleem in plaats van een stochastisch spel

Een markov-strategie voor speler 2 voor dit t-staps spel noteren we als π_2^t en iedere π_2^t is een rij afbeeldingen $\{\pi_2^t(\tau)\}_{\tau=1}^{t}$ met $\pi_2^t(\tau) : Q \to \Pi_2$, d.w.z. $\pi_2^t(\tau)$ voegt aan iedere $\bar{q} \in Q$ een $\bar{\pi} \in \Pi_2$ toe (notatie $\pi_2^t(\bar{q}, \tau)$).
Als speler 2 de strategie π_2^t speelt en als op tijdstip τ toestand $\bar{q}$ optreedt, dan voert speler 2 de gemengde aktie $\pi_2^t(\bar{q}, \tau)$ uit.
De totale verwachte kosten voor speler 2 bij strategie π_2^t noteren we als $W_2^t(\bar{q}, \pi_2^t)$ als $\bar{q}$ de begintoestand van het systeem is.

Laat $W_t^*(\bar{q}) = \min_{\pi_2^t} W_2^t(\bar{q},\pi_2^t)$.

Het enige verschil tussen het t-staps spel zoals boven geformuleerd en het t-staps spel zoals behandeld in sektie 1.2 is, dat we hier te maken hebben met een overaftelbare toestandsruimte S, terwijl die in sektie 1.2 eindig is.

Echter mede door de eenvoudige struktuur van de overgangskansen gaat stelling 1.2.1 ook voor bovenstaand model op.

Gevolg:

(2.1) $$W_t^*(\bar{q}) = \min_{\bar{\pi}} \sum_{i=1}^{N} \pi(i) \{ (c_i - p_i q_i r_i) + (1-p_i q_i) W_t^*(\bar{q}^i) \}$$

waarbij $\bar{q}^i = (q_1^i,\ldots,q_N^i)$, $q_i^i = \dfrac{(1-p_i)q_i}{1-p_i q_i}$ en $q_j^i = \dfrac{q_j}{1-p_i q_i}$, $j \neq i$.

Uit (2.1) blijkt, dat het minimum aangenomen wordt door een zuivere aktie $\bar{\pi}$. Een optimale markov-strategie voor speler 2 kan op dezelfde wijze gevonden worden als in stelling 1.2.1 aangegeven.

Opmerking. Voor een specifiek spel met begintoestand $\bar{q}^0$ kan het probleem worden gezien als een eindige toestanden probleem. Het aantal toestanden neemt echter wel exponentieel toe als t (het aantal perioden) groter wordt. Het systeem kan zich na de eerste periode namelijk slechts in één van de toestanden $\bar{q}^{01},\ldots,\bar{q}^{0N}$ bevinden, waarbij

$$q_i^{0i} = \frac{(1-p_i)q_i^0}{1-p_i q_i^0} \quad \text{en} \quad q_j^{0i} = \frac{q_j^0}{1-p_i q_i^0} \qquad j \neq i,\ j=1,\ldots,N;\ i=1,\ldots,N.$$

Aldus doorgaande blijkt, dat in t perioden hoogstens N^t+1 toestanden bereikt kunnen worden. We zouden dan het probleem met vaste begintoestand kunnen oplossen door alleen van deze N^t+1 toestanden gebruik te maken.

We beschouwen nu een probleem, waarbij de doelstelling voor speler 2 van het type 4 is.

Bijvoorbeeld, maximaliseer de verwachte vindkans onder de conditie dat de verwachte kosten niet groter zijn dan $B_t(\bar{q})$.

Uiteraard moet gelden $B_t(\bar{q}) \geq W_t(\bar{q})$.

Dit probleem lossen we op door de toestandsruimte uit te breiden tot

$\{Q \times [0,B_t(\bar{q})] \cup \{0,1\}\}$

De overgangskansen bij een gemengde aktie $\bar{\pi} \in \Pi_2$ in toestand $(\bar{q},b)$ zijn nu:

met kans $\sum_{i=1}^{N} \pi(i) q_i p_i$ overgaan naar toestand 1 en voor $i=1,\dots,N$:

met kans $\pi(i)(1-p_iq_i)$ overgaan naar toestand $\left(\bar{q}^i, b-\frac{\pi(i)(c_i-p_iq_ir_i)}{\pi(i)(1-p_iq_i)}\right)$

waarbij $\bar{q}^i$ gedefinieerd is als in (2.1).

Merk op dat iedere gemengde aktie uit het vorige model nu opgevat moet worden als een zuivere aktie.

De toestanden $(\bar{q},b)$, waarin $b< \min_i (c_i-p_iq_ir_i)$ kunnen met de stoptoestand 0 geïdentificeerd worden.

De aktieruimte in toestand $(\bar{q},b)$ is:

$$\Pi_2(\bar{q},b)= \{\bar{\pi} \mid \pi\in\Pi_2,\ \sum_{i=1}^{N} (c_i-p_iq_ir_i)\pi(i) \geq b \} . \tag{2.2}$$

De direkte verwachte vindkans voor speler 2 als hij $\bar{\pi}\in\Pi_2(\bar{q},b)$ in tostand $(\bar{q},b)$ speelt is $\sum_{i=1}^{N} \pi(i)p_iq_i$.

Laat $P_t^*(\bar{q},B_t(\bar{q}))$ de maximale verwachte vindkans voor speler 2 voor een t-staps spel met starttoestand $(\bar{q},B_t(\bar{q}))$ zijn.

Passen we nu weer stelling 1.2.1 toe dan vinden we:

$$P_t^*(\bar{q},B_t(\bar{q}))=\max_{\bar{\pi}\in\Pi_2(\bar{q},b)} \sum_{i=1}^{N} \pi(i) \{p_iq_i+(1-p_iq_i). \tag{2.3}$$
$$.P_{t-1}^*\left(\bar{q}^i,B_t(\bar{q})-\frac{\pi(i)(c_i-p_iq_ir_i)}{\pi(i)(1-p_iq_i)}\right)\}$$

Hierbij is $P_{t-1}^*(\bar{q},b)=0$ voor $b < \min_i (c_i-p_iq_ir_i)$.

Uit (2.3) blijkt nu, dat het maximum niet meer door een zuivere aktie aangenomen hoeft te worden, zodat ook de optimale markov-strategie in het algemeen gemengd zal zijn.

Voor bekende $P_{t-1}^*(.,.)$ vormen (2.3) en (2.2) een lineair programmeringsprobleem.

Dit laatste model kunnen we ook beschouwen voor het geval, dat de gerealiseerde kosten in een t-staps spel een bepaald bedrag $B_t^{\bar{q}}$ niet mogen overschrijden . In dat geval is

$$\Pi_2(\bar{q},b)= \{\bar{\pi} \mid \bar{\pi}\in\Pi_2,\ \pi(i)=0 \text{ als } c_i>b,\ i=1,\dots,N\} .$$

Het valt gemakkelijk in te zien dat er nu wel weer een zuivere $\bar{\pi}$ is, waarvoor het maximum in (2.3) aangenomen wordt.

Model 1^b. De enige verandering t.o.v. model 1^a is, dat we nu te maken hebben met een oneindige horizon.
De opbrengsten worden verdisconteerd m.b.v. een factor β, $0 \le \beta < 1$.
Een markov-strategie π_2^M voor speler 2 is nu een oneindige rij afbeeldingen

$\{\pi_2^M(\tau)\}_{\tau=1}^{\infty}$ met $\pi_2^M(\tau) : Q \to \Pi_2$,alle τ. (notatie $\pi_2^M(\tau,\bar{q})$)

Laat $W(\bar{q},\pi_2^M)$ de totale verwachte verdisconteerde kosten zijn voor speler 2 als hij strategie π_2^M speelt en de begintoestand $\bar{q}$ is.
Definieer bij een markov-strategie π_2^M de markov-strategie π_2^{M-1} als
$\pi_2^{M-1}(\tau) = \pi_2^M(\tau+1)$, $\tau=1,2,\ldots$
Eenvoudig is na te gaan, dat geldt:

$$W(\bar{q},\pi_2^M) = \sum_{i=1}^{N} \pi_2^M(1,\bar{q},i) \,.\{c_i - p_i q_i r_i + \beta(1-p_i q_i) W(\bar{q}^i,\pi_2^{M-1})\} ,$$

met $\bar{q}^i$ gedefinieerd zoals in (2.1). Laat $W^*(\bar{q}) = \min\limits_{\pi_2^M} W(\bar{q},\pi_2^M)$.

Evenals in sektie 1.2, stelling 1.2.2 kan ook nu (overaftelbare toestandsruimte) aangetoond worden, dat $W^*(\bar{q})$ de unieke oplossing is van de functionaalvergelijking:

(2.4) $$W^*(\bar{q}) = \min_{\bar{\pi}} \sum_{i=1}^{N} \pi(i) \{c_i - p_i q_i r_i + \beta(1-p_i q_i) W^*(\bar{q}^i)\}$$

De vergelijking (2.4) wordt wel de optimaliteitsvergelijking van Bellman [2] genoemd.
Met behulp van (2.4) is tevens een optimale zuivere stationaire strategie te bepalen op identieke wijze als in stelling 1.2.3.; in toestand $\bar{q}$ kiest speler 2 die zuivere $i \in \{1,\ldots,N\}$, waarvoor in (2.4) het minimum aangenomen wordt.
Benaderingen van de waarde en van de optimale strategieën kunnen gevonden worden zoals in stelling 1.2.5. aangegeven met behulp van de methode der successieve approximatie.

Uitbreidingen van dit model naar modellen met doelstellingen van het type 4 verlopen soortgelijk als onder model 1^a.
Bijvoorbeeld, maximaliseer de totale verwachte verdisconteerde pakkans onder de conditie, dat de totale verwachte verdisconteerde kosten beneden een bepaalde waarde $B(\bar{q})$ blijven.
Laat $P^*(\bar{q},B(\bar{q}))$ deze maximale verwachte verdisconteerde pakkans zijn.

Op de bekende manier volgt weer dat $P^*(\bar{q},B(\bar{q}))$ de unieke oplossing is van de optimaliteitsvergelijking:

$$(2.5)\qquad P^*(\bar{q},B(\bar{q}))=\max_{\bar{\pi}\in\Pi_2(B(\bar{q}))}\sum_{i=1}^{N}\pi(i)\ \{p_iq_i+\beta(1-p_iq_i)\ .\ P^*\left(\bar{q}^i,\frac{B(\bar{q})-c_i+p_iq_ir_i}{\beta(1-p_iq_i)}\right)\},$$

waarbij $\Pi_2(B(\bar{q}))=\{\bar{\pi}\mid\bar{\pi}\in\Pi_2,\ \sum_{i=1}^{N}\pi_i(c_i-p_iq_ir_i)\ \geq B(\bar{q})\}$

en $\bar{q}^i$ zoals gedefinieerd in (2.1).

Uit (2.5) blijkt, dat de optimale stationaire strategie voor speler 2 niet meer zuiver hoeft te zijn.

We merken nog op, dat $P^*(\bar{q},B(\bar{q}))$ voor alle $\bar{q}$ continu is in $B(\bar{q})$ waarvoor $B(\bar{q})\geq\min_i c_i-p_iq_ir_i$ en voor alle $B(\bar{q})\geq\min_i c_i-p_iq_ir_i$ continu is in $\bar{q}$ en tevens $P^*(\bar{q},B(\bar{q}))=0$ voor $B(\bar{q})<\min_i c_i-p_iq_ir_i$ en alle $\bar{q}$.

Model 2[a]. We beperken ons tot het model, waarbij speler 1 zich maar op één manier in ieder doosje kan verstoppen en speler 2 maar op één manier in een doosje kan kijken. Uitbreidingen zijn zonder moeilijkheid mogelijk.

Toestandsruimte: $\{1,\ldots,N\}\times\{1,\ldots,N\}\cup\{0,1\}$.

Hierbij is 0 weer de stoptoestand en niet gevonden, 1 de stoptoestand en wel gevonden en (k_1,k_2) is de toestand, die overeen komt met de situatie: speler 1 heeft zich in de vorige periode in doos k_1 verstopt en speler 2 heeft in doos k_2 gezocht.

In dit model gaan we er dus vanuit, dat aan het einde van iedere periode de beide spelers zowel hun eigen positie als de positie van de andere speler weten.

Aktieruimte voor speler n in toestand (k_1,k_2): $A_n(k_1,k_2)\subset\{1,\ldots,N\}$.

Het kan zijn dat een speler in een periode niet van iedere doos naar iedere andere doos kan komen.

Laat $\Pi=\{\bar{\pi}\mid\bar{\pi}=(\pi(1),\ldots,\pi(N)),\ \pi(i)\geq 0,\ \sum_{i=1}^{N}\pi(i)=1\}$.

De gemengde aktieruimte voor speler n in toestand (k_1,k_2) is dan :

$\Pi_n(k_1,k_2)=\{\bar{\pi}_n\mid\bar{\pi}_n\in\Pi,\ \pi_n(i)=0$ als $i\notin A_n(k_1,k_2)\}$.

Als in toestand (k_1,k_2) de spelers zuivere akties $i\in A_1(k_1,k_2)$ en $j\in A_2(k_1,k_2)$ kiezen, dan horen hierbij de volgende overgangskansen:

$$p((l_1,l_2) \mid (k_1,k_2),i,j) = 1 \text{ voor } (l_1,l_2)=(i,j),\ i\neq j$$
$$= 0 \text{ voor } (l_1,l_2)\neq(i,j),\ i\neq j$$
$$= p_i \text{ voor } (l_1,l_2)=(i,j),\ i=j$$
$$= 0 \text{ voor } (l_1,l_2)\neq(i,j),\ i=j$$
$$p(1 \mid (k_1,k_2),i,j) = 1-p_i \text{ voor } i=j$$
$$= 0 \text{ voor } i\neq j$$

Als we te maken hebben met een t-staps probleem dan gelden de bovenstaande overgangskansen voor $\tau < t$. Voor periode t geldt:

$$p(1 \mid (k_1,k_2),i,j) = p_i \text{ voor } i=j$$
$$= 0 \text{ voor } i\neq j$$
$$p(0 \mid (k_1,k_2),i,j) = 1-p_i \text{ voor } i=j$$
$$= 1 \text{ voor } i\neq j$$

De direkte verwachte kosten voor speler 2 in toestand (k_1,k_2) als de spelers $\overline{\pi}_1\in\Pi_1(k_1,k_2)$ en $\overline{\pi}_2\in\Pi_2(k_1,k_2)$ spelen is:

$$\sum_{i=1}^{N} \sum_{j=1}^{N} \pi_1(i)\pi_2(j)\ (c_j-p_jr_j\delta_{ij}) ,$$

waarbij $\delta_{ij}=1$ als $i=j$ en $\delta_{ij}=0$ als $i\neq j$.
Markov-strategieën kunnen op dezelfde manier gedefinieerd worden als in de vorige modellen.
Model 2^a, zoals hierboven geformuleerd is exact een voorbeeld van een t-staps spel zoals behandeld in hoofdstuk 1. Nemen we als doelstelling de kosten, dan bezit dit spel een waarde en hebben beide spelers optimale markov-strategieën, die gevonden kunnen worden op de manier zoals in hoofdstuk 1 stelling 1.2.1. aangegeven.

Ook nu kunnen we doelstellingen van het type 4 beschouwen. Bijvoorbeeld, speler 2 optimaliseert zijn vindkans terwijl hem een bedrag B ter beschikking staat om zijn kosten te dekken; meer kosten maken dan B is verboden.
We kunnen het bedrag, dat speler 2 nog ter beschikking staat in de toestand opnemen. Aannemende, dat zowel B als alle c_i's, $i=1,\dots,N$ natuurlijke getallen zijn, kunnen we nu als toestandsruimte nemen:

$\{1,\dots,N\} \times \{1,\dots,N\} \times \{1,\dots,B\} \cup \{0,1\}$, waarbij (k_1,k_2,b) nu

betekent, dat speler 1 zich in de vorige periode in doos k_1 verstopt heeft, dat speler 2 tevergeefs in doos k_2 gezocht heeft en dat speler 2 nu nog een bedrag b ter beschikking staat.
De aktieruimte voor speler 2 moet aangepast worden:

$$A_2(k_1,k_2,b)= \{j \mid j\in A_2(k_1,k_2);\ c_j \geq b\} .$$

Voor speler 1 blijft de aktieruimte ongewijzigd.
De direkte opbrengst bij zuivere akties (i,j) in toestand (k_1,k_2,b) is nu p_j als i=j en 0 als i≠j.
De overgangskansen ondergaan een evidente wijziging:

$$p((l_1,l_2,b') \mid (k_1,k_2,b),i,j)= p_j \text{ voor } i=j,\ (l_1,l_2)=(i,j),\ b'=b-c_j$$

enz.

Hiermede is wederom een eindigstaps stochastisch spel geformuleerd, zoals behandeld in hoofdstuk 1.

Model 2^b. De stochastische spelformulering van dit model is gelijk aan de formulering van model 2^a, met dien verstande, dat de overgangskansen dezelfde zijn voor iedere periode (in model 2^a iets afwijkends voor de laatste periode) en dat de toekomstige kosten (of vindkansen) verdisconteerd worden met behulp van een verdisconteringsfactor β, $0\leq\beta<1$. Het spel stopt alleen als de verstopper gevonden wordt.

Beschouwen we het probleem van het optimaliseren van de verwachte kosten, dan verkrijgen we weer een spel zoals behandeld in hoofdstuk 1. Uit de stellingen 1.2.2. en 1.2.3. volgt weer het bestaan van de waarde en optimale stationaire strategieën voor de beide spelers.

We beschouwen nu het probleem van de verwachte duur van het spel. De toestandsruimte is $S= \{1,\ldots,N\} \times \{1,\ldots,N\} \cup \{0,1\}$.
Een stationaire strategie voor speler n is een afbeelding $\pi_n : S \to \Pi$, z.d.d. $\pi_n(k_1,k_2)\in\Pi_n(k_1,k_2)$.
Laat $D((k_1,k_2),\pi_1,\pi_2)$ de verwachte duur van het spel, als de spelers de stationaire trategieën π_1 en π_2 spelen en (k_1,k_2) de begintoestand is.

Aanname: De zuivere aktieruimten in de diverse toestanden voor de beide spelers zijn dusdanig, dat onder ieder paar zuivere stationaire strategieën de bijbehorende Markov-keten absorberend is met 1 als absorberende toestand en de toestanden (k_1,k_2), $k_1,k_2=1,\ldots,N$ als doorgangstoestanden.

Uit deze aanname kunnen we afleiden, dat voor ieder paar stationaire strategieën π_1 en π_2 de bijbehorende Markov-keten absorberend is en tevens, dat er een T en een $\varepsilon>0$ bestaat, z.d.d. voor de T-stapsovergangskansen geldt:

(2.6) $$\sum_{l_1=1}^{N} \sum_{l_2=1}^{N} p^T((l_1,l_2) \mid (k_1,k_2),\pi_1,\pi_2) \le 1-\varepsilon ,$$

alle $k_1,k_2=1,\ldots,N$, alle π_1 en π_2.

Analoog aan (1.5) geldt nu:

$$\begin{aligned} D((k_1,k_2),\pi_1,\pi_2) &= 1 \cdot \sum_{i=1}^{N} \pi_1((k_1,k_2),i)\pi_2((k_1,k_2),i)p_i + \\ &+ \sum_{i=1}^{N} \sum_{j=1}^{N} \pi_1((k_1,k_2),i)\pi_2((k_1,k_2),j) . \\ &\cdot \sum_{l_1=1}^{N} \sum_{l_2=1}^{N} p((l_1,l_2) \mid (k_1,k_2),i,j).(1+D((l_1,l_2),\pi_1,\pi_2)) \\ &= \sum_{i=1}^{N} \pi_1((k_1,k_2),i)\pi_2((k_1,k_2),i)p_i \\ &+ \sum_{i=1}^{N} \sum_{j=1}^{N} \pi_1((k_1,k_2),i)\pi_2((k_1,k_2),j)(1-p_j\delta_{ij}) . \\ &\cdot (1+D((i,j),\pi_1,\pi_2)) \\ &= 1+ \sum_{i=1}^{N} \sum_{j=1}^{N} \pi_1((k_1,k_2),i)\pi_2((k_1,k_2),j) . \\ &\cdot (1-p_j\delta_{ij})\, D((i,j),\pi_1,\pi_2) \quad , \end{aligned}$$

zodat

(2.7) $$\begin{aligned} D((k_1,k_2),\pi_1,\pi_2) = 1+ &\sum_{i=1}^{N} \sum_{j=1}^{N} \pi_1((k_1,k_2),i)\pi_2((k_1,k_2),j) . \\ &\cdot \{(1-p_j\delta_{ij})D((i,j),\pi_1,\pi_2)+p_j\delta_{ij}D(1,\pi_1,\pi_2)\} \end{aligned}$$

Hierbij is $D(1,\pi_1,\pi_2)= 0$, want gevonden blijft gevonden.

Uit (2.7)blijkt nu, dat $D((k_1,k_2),\pi_1,\pi_2)$ gelijk is aan de totale verwachte uitbetaling in een stochastisch spel, waarbij de direkte uitbetaling 1 is in iedere toestand bij ieder paar akties en de andere componenten zoals bovenstaand geformuleerd.

De aanname, die resulteert in de eigenschap (2.6) is een ruimere voorwaarde,dan de voorwaarde 'hetspel is stoppend'.

In het laatste geval verkrijgen we (2.6) voor T=1. Onder de voorwaarde (2.6) hebben we te maken met een T-staps contractie en het valt op dezelfde manier als voor T=1 (hoofdstuk 1, sektie 1.2) aan te tonen, dat

ook nu de totale verwachte uitbetaling eindig blijft.
Evenzo blijven de stellingen 1.2.2. en 1.2.3. geldig. De waarde en optimale strategieën voor het probleem van de verwachte duur van het spel volgen nu uit de oplossing van het verwachte kosten spel, door als direkte betaling 1 te nemen in iedere toestand bij ieder paar akties.

We gaan nu weer een doelstelling van het type 4 beschouwen voor speler 2: optimaliseer de verwachte kosten onder de conditie, dat de verwachte duur van het spel niet groter is dan een bepaalde waarde.
Nummer de toestanden, z.d.d. de stoptoestand 1 de laatste is.
Laat $P^{t0}_{\pi_1\pi_2}$ de matrix, die overblijft door in de t-stapsovergangskansen-matrix $P^{t}_{\pi_1\pi_2}$ de laatste kolom en de laatste rij te schrappen, $t=1,2,\ldots$.
Laat $\bar{e}_0$ een kolomvektor zijn, die evenveel komponenten heeft als er toestanden zijn en die bestaat uit allemaal nullen behalve de laatste komponent, die de waarde 1 heeft. Laat $\bar{e}_1$ een kolomvektor zijn, die bestaat uit enen en die evenveel komponenten heeft als er toestanden (k_1,k_2) zijn.
Laat $\bar{D}(\pi_1,\pi_2)$ de vektor zijn met komponenten $D((k_1,k_2),\pi_1,\pi_2)$.
Dan geldt:

$$(2.8)\qquad \begin{aligned}\bar{D}(\pi_1,\pi_2) &= \sum_{t=1}^{T} t(P^{t}_{\pi_1\pi_2}-P^{t-1}_{\pi_1\pi_2}).\bar{e}_0 + P^{T0}_{\pi_1\pi_2}(T.\bar{e}_1+\bar{D}(\pi_1,\pi_2))\\ &= T.P^{T}_{\pi_1\pi_2}.\bar{e}_0 - \sum_{t=1}^{T-1} P^{t}_{\pi_1\pi_2}.\bar{e}_0 + P^{T0}_{\pi_1\pi_2}(T.\bar{e}_1+\bar{D}(\pi_1,\pi_2))\\ &\le T\{P^{T}_{\pi_1\pi_2}.\bar{e}_0 + P^{T0}_{\pi_1\pi_2}.\bar{e}_1\} + P^{T0}_{\pi_1\pi_2}.\bar{D}(\pi_1,\pi_2)\\ &= T.\bar{e}_1 + P^{T0}_{\pi_1\pi_2}.\bar{D}(\pi_1,\pi_2)\end{aligned}$$

Hierbij is $P^{0}_{\pi_1\pi_2}$ een matrix bestaande uit allemaal nullen.

Laat $D((k_1^0,k_2^0),\pi_1,\pi_2) = \max_{(k_1,k_2)} D((k_1,k_2),\pi_1,\pi_2)$.
Uit (2.8) volgt:

$$D((k_1^0,k_2^0),\pi_1,\pi_2) \le T + D((k_1^0,k_2^0),\pi_1,\pi_2).P^{T0}_{\pi_1\pi_2}(k_1^0,k_2^0).\bar{e}_1 \qquad \text{of}$$

$$(2.9)\qquad D((k_1^0,k_2^0),\pi_1,\pi_2) \le \frac{T}{1 - P^{T0}_{\pi_1\pi_2}(k_1^0,k_2^0).\bar{e}_1} \quad ,$$

hierbij is $P^{TO}_{\pi_1\pi_2}(k_1^0,k_2^0)$ de $(k_1^0,k_2^0)^e$ rij van de matrix $P^{TO}_{\pi_1\pi_2}$.
Als speler 2 zich nu tot de stationaire strategieën beperkt, waarvoor:

(2.10) $$\max_{\pi_1} \max_{(k_1,k_2)} P^{TO}_{\pi_1\pi_2}(k_1,k_2) \le (1-\frac{1}{D}),$$

dan blijkt uit (2.9):

$$D((k_1,k_2),\pi_1,\pi_2) \le D((k_1^0,k_2^0),\pi_1,\pi_2) \le \frac{T}{1-(1-\frac{T}{D})} = D, \text{ alle } \pi_1 \text{ en } (k_1,k_2),$$

waarmee een begrenzing op de verwachte duur van het spel is aangegeven.
Als T=1 geeft dit lineaire nevenvoorwaarden, namelijk:

(2.11) $$\sum_{j=1}^{N} p_j\delta_{ij}\pi_2((k_1,k_2),j) \ge 1-\frac{1}{D} \quad , \text{ alle } i\in A_1(k_1,k_2), \text{ alle } \pi_2.$$

Eenvoudig is in te zien, dat dit alleen maar op kan treden als beide spelers in iedere toestand (k_1,k_2) slechts één en voor beiden dezelfde doos kunnen kiezen. Wat overblijft van de voorwaarde (2.11) is dan $p_j \ge 1-\frac{1}{D}$ en dit legt geen beperking op de strategieruimte voor speler 2, daar p_j een gegeven spelparameter is.
Deze nevenvoorwaarde is niet triviaal als speler 2 voor de betreffende doos meerdere zoekmogelijkheden heeft, bijvoorbeeld 1,...,m .
De vindkans voor deze doos is dan afhankelijk van de gekozen aktie.
Een gemengde aktie voor speler 2 in toestand (k_1,k_2) is dan een vektor $(\pi_2^1(k_1,k_2),\ldots,\pi_2^m(k_1,k_2))$ met $\pi_2^j(k_1,k_2)\ge 0$ en $\sum_{j=1}^{m}\pi_2^j(k_1,k_2)=1$.
Voorwaarde (2.11) wordt nu:

(2.12) $$\sum_{j=1}^{m} \pi_2^j(k_1,k_2)\cdot p_{ij}(k_1,k_2) \ge 1-\frac{1}{D} \quad , \text{ alle } i \ ,$$

waarbij $p_{ij}(k_1,k_2)$ de vindkans in toestand (k_1,k_2) is als speler 2 alternatief j kiest en speler 1 alternatief i.
Voorwaarde (2.12) is niet triviaal voor $p^0 \ge (1-\frac{1}{D}) \ge p_0$,
waarbij $p_0 = \min_{ijk_1k_2} p_{ij}(k_1,k_2)$ en $p^0 = \max_{ijk_1k_2} p_{ij}(k_1,k_2)$.
Voor T>1 is de uitdrukking (2.10) in het algemeen niet triviaal en hebben we te maken met niet-lineaire nevenvoorwaarden.

3.4 UITWERKEN VAN VOORBEELDEN.

In deze sektie gaan we aangeven, hoe de voorbeelden van sektie 3.1 m.b.v. de modellen uit sektie 3.3 geanalyseerd kunnen worden. Tevens wordt voor het eerste probleem een eenvoudig geval uitgewerkt.

a. Het surveilleer probleem.

De politie is de zoeker en de bende is de verstopper. We veronderstellen, dat het probleem zich uitstrekt over een tijdvak van t nachten.
De bende kan zich iedere nacht ophouden in één van N banken (dozen) of zich schuil houden, hetgeen neerkomt op verstoppen in een $(N+1)^e$ doos.
We veronderstellen, dat de politie per nacht slechts één bank kan bewaken. (Indien meerdere banken bewaakt zouden kunnen worden, dan moet de politie als een eenheid worden gezien, die in te kiezen porties over te kiezen banken verdeeld kunnen worden.)
We hebben te maken met een t-staps zoekprobleem, waarbij beide spelers iedere nacht opnieuw een keuze mogen doen.
Model 2^a is dus van toepassing.
Pakkans p_i en opbrengst r_i voor een geslaagde kraak op bank i zijn bankafhankelijk.

De bende kan zich bijvoorbeeld ten doel stellen hun verwachte opbrengst te maximaliseren onder de conditie, dat de maximaal te verwachten pakkans kleiner is dan een van te voren aanvaard risico p_0.
Als toestandsruimte moeten we dan nemen $\{[0,p_0]\} \cup \{0,1\}$.
0 en 1 hebben de betekenis als in de vorige sektie. Toestand $q \in [0,p_0]$ betekent, dat de inbrekers in de resterende stappen nog een maximale verwachte pakkans q kunnen aanvaarden.
Laat $V^t(p_0)$ de waarde van het spel zijn voor begintoestand p_0.
Toepassen van stelling 1.2.1. levert:

$$(2.13) \qquad V^t(p_0) = \underline{\max}_{\pi_1 \in \Pi_1(p_0)} \; \underline{\min}_{\pi_2} \; \sum_{i=1}^{N} \pi_1(i)(1-\pi_2(i))r_i + \Big(1 - \sum_{i=1}^{N}\pi_1(i)\pi_2(i)\Big) .$$

$$. \, V^{t-1}\left(\frac{p_0 - \sum_{i=1}^{N}\pi_1(i)\pi_2(i)}{1 - \sum_{i=1}^{N}\pi_1(i)\pi_2(i)} \right) ,$$

hierbij is $\Pi_1(p_0)=\{\ \bar{\pi}\ |\ \bar{\pi}=(\pi(1),\ldots,\pi(N)),\ \pi(i).p_i \le p_0,\ i=1,\ldots,N\}$.
Eveneens met behulp van stelling 1.2.1. volgen optimale markov-strategieën voor beide spelers.

We gaan nu een vereenvoudigd geval nader uitwerken.
We beschouwen een situatie met 2 banken, 2 politiemensen, 2 inbrekers en 2 nachten waarin ingebroken kan worden.
De inbrekers werken gezamenlijk, maar de politiemensen kunnen ook afzonderlijk ieder een bank bewaken.
De inbrekers hebben dan per nacht twee zuivere akties en de politiemensen drie. Dit is terug te vinden in onderstaande tabellen, waarin de pakkansen en de verwachte opbrengst voor iedere combinatie van zuivere akties getabelleerd staan.

		politie (2,0)	(1,1)	(0,2)
inbrekers	1	1	$\frac{1}{2}$	0
	2	0	$\frac{1}{2}$	1

Tabel voor pakkansen

		politie (2,0)	(1,1)	(0,2)
inbrekers	1	0	$\frac{1}{2}r_1$	r_1
	2	r_2	$\frac{1}{2}r_2$	0

Tabel voor verwachte opbrengst

Aktie (i,j) voor de politiemensen betekent i agenten bij bank 1 en j agenten bij bank 2.
Merk op, dat de pakkans als de inbrekers bank i kraken gelijk is aan $\frac{n_P(i)}{n_B(i)}$ met $n_P(i)$ het aantal agenten bij bank i en $n_B(i)$ het aantal inbrekers bij bank i.
Uit de beide tabellen is te zien, dat de zuivere aktie (1,1) voor de politie dezelfde verwachte uitkomsten geeft als de gemengde aktie 'met kans $\frac{1}{2}$ aktie (2,0) en met kans $\frac{1}{2}$ aktie (0,2)'.
Wat de verwachte uitkomsten betreft, verliest de politie dus niets als ze hun gemengde akties uitsluitend baseren op (2,0) en (0,2).
In de navolgende berekeningen wordt dit dan ook gedaan.

Als doelstelling voor de inbrekers nemen we het maximaliseren van de verwachte opbrengst over deze twee nachten, onder de conditie, dat de verwachte pakkans niet groter is dan 5/6.
Als toestandsruimte moeten we nemen: $\{[0,5/6]\} \cup \{0,1\}$.

Uit de tabel voor de pakkansen volgt min max (pakkans)= $\frac{1}{2}$, zodat de toestanden $\{[0,\frac{1}{2})\}$ met de 'stoptoestand en niet gepakt' 0 geïdentificeerd kunnen worden.
De gemengde aktieruimte voor de inbrekers in toestand $q\in[\frac{1}{2},5/6]$ is:

(2.14) $$\Pi_1(q)= \{\bar{\pi} \mid \bar{\pi}=(\pi(1),\pi(2)),\ \pi(1)=1-\pi(2),\ \pi(1)\in[1-q,q]\} .$$

Voor beide spelers geldt, dat een gemengde aktie gekarakteriseerd wordt door het gewicht, dat ze op hun eerste aktie leggen. In het vervolg duiden we gemengde akties daar dan ook mee aan. Voor de politie gebruiken we hiervoor de notatie π_2 en voor de inbrekers π_1.
Toestandsovergangskansen bij π_1, π_2:

op tijdstip 1: met kans $\pi_1\pi_2+(1-\pi_1)(1-\pi_2)$ naar stoptoestand 1

met kans $(1-\pi_1)\pi_2+\pi_1(1-\pi_2)$ naar toestand $\dfrac{5/6-\pi_1\pi_2-(1-\pi_1)(1-\pi_2)}{(1-\pi_1)\pi_2+\pi_1(1-\pi_2)}$

optijdstip 2: met kans $\pi_1\pi_2+(1-\pi_1)(1-\pi_2)$ naar stoptoestand 1

met kans $(1-\pi_1)\pi_2+\pi_1(1-\pi_2)$ naar stoptoestand 0

In ons model is geen nabetalingsfunctie opgenomen. Eigenlijk ten onrechte, daar het voor de inbrekers wel degelijk verschil uitmaakt of ze na afloop in de gevangenis zitten (toestand 1) of van het geld kunnen genieten (toestand 0).

Aanname: $r_1=3r_2$
Laat $V^t(q)$ de waarde van het spel als er t stappen te gaan zijn en de toestand q is.
Dan geldt:

$$V^1(q)= \max_{\pi_1\in[1-q,q]} \min_{\pi_2} \{(1-\pi_1)\pi_2 r_2+\pi_1(1-\pi_2)r_1\} \quad , q\in[\tfrac{1}{2},5/6)$$

We kunnen opmerken, dat dit het zelfde is als het vinden van de waarde van het matrix-spel:

$$\begin{pmatrix} qr_2 & (1-q)r_1 \\ (1-q)r_2 & qr_1 \end{pmatrix} ,$$

dat geïnterpreteerd moet worden als een spel, waarbij de twee zuivere akties voor de politie (2,0) en (0,2) zijn en de twee zuivere akties voor de inbrekers (1-q,q) en (q,1-q) zijn.

Er volgt:

$V^1(q)= qr_2$ voor $q\in[\frac{1}{2},\frac{3}{4}]$, beide eerste akties vormen zadelpunt.

(2.15) $V^1(q)= \frac{3}{4}r_2$ voor $q\in(\frac{3}{4},5/6]$, voor de politie is $(\frac{3}{4},\frac{1}{4})$ optimaal en voor de inbrekers $(\frac{p-\frac{1}{4}}{2p-1}\ ,\ \frac{p-\frac{3}{4}}{2p-1})$.

$V^1(q)= 0$ voor $q\in[0,\frac{1}{2})$.

(2.13) toepassen geeft:

(2.16)
$$V^2(5/6)= \max_{\pi_1\in[1/6,5/6]}\ \min_{\pi_2}\{\pi_1(1-\pi_2)r_1+(1-\pi_1)\pi_2 r_2+ \\ +\ (\pi_1(1-\pi_2)+\pi_2(1-\pi_1)).V^1(\frac{5/6-\pi_1\pi_2-(1-\pi_1)(1-\pi_2)}{\pi_1(1-\pi_2)+\pi_2(1-\pi_1)})$$

We noteren de uitdrukking in (2.16) tussen accoladen als $G(\pi_1,\pi_2)$.
Als $\pi_1\in[1/3,2/3]$, dan $\frac{5/6-\pi_1\pi_2-(1-\pi_1)(1-\pi_2)}{\pi_1(1-\pi_2)+{}_2(1-\pi_1)}\in[\frac{1}{2},\frac{3}{4}]$, alle π_2,

zodat voor $\pi_1\in[1/3,2/3]$ m.b.v. (2.15):

$$\min_{\pi_2} G(\pi_1,\pi_2)= \min_{\pi_2} \pi_2(1-2\pi_1)r_2+(1-\pi_2)(4\pi_2-1)r_2+(5/6)r_2$$

$$= (1-2\pi_1)r_2+(5/6)r_2\ ,$$

waarbij het minimum aangenomen wordt door $\pi_2=1$.
Hieruit volgt:

(2.17)
$$\max_{\pi_1\in[1/3,2/3]}\ \min_{\pi_2} G(\pi_1,\pi_2) =(1-2.1/3)r_2+(5/6)r_2= (7/6)r_2$$

voor $\pi_1=1/3$.

Als $\pi_1\in[1/6,5/6] \setminus [1/3,2/3]$, dan kan de politie het spel in één stap laten stoppen door die zuivere aktie in stap 1 te kiezen, die dezelfde is als de aktie, waarop de inbrekers het meeste gewicht leggen.
Namelijk als $\pi_1\in[1/6,1/3)$, dan kiest de politie de 2^e aktie en geldt voor de volgende toestand $\frac{5/6-(1-\pi_1)}{\pi_1}<\frac{1}{2}$, zodat het spel in een toestand komt, die met de stoptoestand 0 geïdentificeerd is,
en als $\pi_1\in(2/3,5/6]$, dan kiest de politie de eerste aktie en geldt voor de volgende toestand $\frac{5/6-\pi_1}{1-\pi_1}<\frac{1}{2}$, zodat het spel stopt.

Het valt eenvoudig na te gaan, dat de verwachte opbrengst in beide gevallen kleiner dan $(7/6)r_2$ is.

Gevolg:

(2.18) $$\max_{\pi_1 \in [1/6,5/6] \setminus [1/3,2/3]} \min_{\pi_2} G(\pi_1,\pi_2) < (7/6)r_2$$

Kombineren van (2.17) en (2.18) geeft:

(2.19)

(2.19) $$V^2(5/6) = \max_{\pi_1 \in [1/6,5/6]} \min_{\pi_2} g(\pi_1,\pi_2) = (7/6)r_2$$

De unieke optimale strategie voor de inbrekers is:

> (1/3,2/3) in de eerste stap en in de tweede stap (1-p,p) als de toestand $p \in [\frac{1}{2},\frac{3}{4}]$ is.

Als de inbrekers het produkt van de minimale niet-pakkans en de minimale opbrengst een belangrijk getal vinden, dan is dat in dit geval $(1/6).(7/6)r_2 = (7/36)r_2 = (28/144)r_2$

Spelen de inbrekers zodanig, dat hun maximale verwachte pakkans 3/4 is, dan moeten ze in de eerste nacht $(\frac{1}{2},\frac{1}{2})$ spelen (anders is het spel na 1 stap geëindigd) en is hun minimale opbrengst:

$$\min_{\pi_2} \{\tfrac{1}{2}\pi_2 r_2 + \tfrac{1}{2}(1-\pi_2)r_1 + \tfrac{1}{2}V^1(\tfrac{\frac{3}{4}-\frac{1}{2}}{\frac{1}{2}})\} = \tfrac{3}{4}r_2$$

Ook in de tweede stap moeten de inbrekers $(\frac{1}{2},\frac{1}{2})$ spelen, daar $\Pi_1(\frac{1}{2}) = (\frac{1}{2},\frac{1}{2})$ en de toestand na de eerste stap $\frac{1}{2}$ is voor alle π_2 in de eerste stap. Het produkt van de minimale niet-pakkans en de minimale opbrengst is nu $\frac{1}{4}.\frac{3}{4}r_2 = (3/16)r_2 = (27/144)r_2$ en dus kleiner dan in het voorgaande geval.

Indien de inbrekers hun opbrengst gaan maximaliseren zonder zich om de pakkans te bekommeren, dan volgt de waarde aldus:
In de laatste stap kunnen de inbrekers zich $\frac{3}{4}r_2$ garanderen (zie (2.15)).
Dan verkrijgen we:

$$V^2 = \max_{\pi_1} \min_{\pi_2} \pi_1(1-\pi_2)r_1 + (1-\pi_1)\pi_2 r_2 + (\pi_1(1-\pi_2) + \pi_2(1-\pi_1))\tfrac{3}{4}r_2$$
$$= (105/88)r_2$$

De optimale strategie voor de inbrekers is nu (15/22,7/22) in de eerste stap en (1/4,3/4) in de tweede stap.

b. Het medisch probleem.

Het medisch probleem is een voorbeeld van een probleem, dat met behulp van model 1^a geanalyseerd kan worden.
De natuur is de verstopper en de onderzoeker is de zoeker.
Laat $Q = \{\bar{q} \mid \bar{q} = (q_1,\ldots,q_N),\ q_i \geq 0,\ \sum_{i=1}^{N} q_i = 1\}$ zijn.
Begintoestand is beginverstopverdeling $\bar{x} = (x_1,\ldots,x_N)$ en budget B.
Het bedrag B en alle bedragen $c_i(j)$ achten we natuurlijke getallen te zijn.
Als $c = \min_i c_i(j)$, dan is de duur van het spel hoogstens entier $[\frac{B}{c}]$ experimenten.
Als toestandsruimte moeten we nemen $\{q \times \{0,1,\ldots,B\}\} \cup \{0,1\}$, waarbij 0 wederom de stoptoestand en niet gevonden is, 1 de stoptoestand en wel gevonden en $(\bar{q},b)$ de toestand is waarin de onderzoeker nog een budget b heeft en $\bar{q}$ zijn conditionele verstopkansverdeling is.
Zuivere aktieruimte in toestand $(\bar{q},b)$ is $A(\bar{q},b) = \{c_i(j) \mid c_i(j) \geq b\}$.
Toestanden $(\bar{q},b)$ met $b>c$ worden met de stoptoestand 0 geïdentificeerd.
Als in toestand $(\bar{q},b)$ zuivere aktie $c_i(j)$ gekozen wordt, dan is er de direkte opbrengst $q_i p_i(c_i(j))$ en zijn de overgangskansen:

met kans $q_i p_i(c_i(j))$ naar toestand 1 springen

met kans $1-q_i p_i(c_i(j))$ naar toestand $(\bar{q}',b-c_i(j))$ springen,

waarbij $q_i' = \dfrac{(1-p_i(c_i(j)))q_i}{1-q_i p_i(c_i(j))}$ en $q_k' = \dfrac{q_k}{1-q_i p_i(c_i(i))}$, $k \neq i$.

De waarde van het spel en optimale markov-strategieën volgen op de manier, zoals in sektie 2.3 onder model 1^a aangegeven.

We geven nog een voorbeeld van een probleem, dat met behulp van zoekmodellen opgelost kan worden.
Dit voorbeeld ligt in de sfeer van de militaire toepassingen.

We denken ons de situatie in, dat een torpedoboot jaagt op een onderzeeër. Het "spel" begint op het moment, dat ze voor de eerste keer contact hebben, d.w.z. de wederzijdse informatie van elkaars aanwezigheid. Het gebied waar het gebeuren plaats vindt denken we ons opgedeeld in een eindig aantal (N) deelgebieden, waarbinnen we veronderstellen, dat de akties constant blijven. De randen van het gebied

worden opgedeeld in een stuk waar de onderzeeër veilig is (bijvoorbeeld thuishaven) en een deel waar de onderzeeër reddeloos verloren is.
We nemen als voorbeeld hier het eenvoudigste geval. Verfijningen en uitbreidingen in velerlei richtingen zijn denkbaar.
De toestandsruimte is $\{1,\ldots,N\} \times \{1,\ldots,N\} \cup \{0,1\}$.
0 en 1 hebben weer de bekende betekenis en (k_1,k_2) betekent onderzeeër bevindt zich in gebied k_1 en torpedoboot in gebied k_2.
De zuivere aktieruimte voor de beide spelers veronderstellen we in iedere toestand eindig. Dezen kunnen o.a. bestaan uit een gebied, waarheen ze de komende periode gaan varen en uit het kiezen van technische grootheden, zoals snelheid, diepte, aanvalswapen, enz. De trefkans (pakkans) is alleen dan positief als beiden zich in het zelfde gebied bevinden en hangt af van de keuze van de technische grootheden.
Een stationaire strategie voor een speler is het aan ieder paar (k_1,k_2) toevoegen van een gemengde aktie.
Als we veronderstellen, dat onder ieder paar stationaire strategieën het spel met kans 1 in één van de beide stoptoestanden komt, dan hebben we een voorbeeld van een model, waarin de resultaten uit sektie 2.3 onder model 2^b direkt toegepast kunnen worden.

2.5 LITERATUUROVERZICHT.

In deze sektie zullen we een beknopt overzicht geven van enkele artikelen, die handelen over zoekproblemen.

a. Neuts (1963) [11].

Neuts behandelt twee modellen. Het eerste model is precies hetzelfde als model 2^b, waarbij ieder doosje maar op één manier geïnspecteerd kan worden. De zoeker minimaliseert zijn kosten.
Toekomstige uitgaven worden verdisconteerd.
Neuts geeft analytische uitdrukkingen voor de waarde en de optimale strategieën van de beide spelers.
Het tweede model, dat Neuts behandelt komt overeen met model 1^b.
Ook hier werkt hij met een verdisconteringsfactor. Bestaan en uniciteit van de oplossing worden via de optimaliteitsvergelijking van Bellman aangetoond.
Ten slotte geeft Neuts nog enkele resultaten voor dit laatste model, waarbij hij de verdisconteringsfactor naar 1 laat gaan

b. Charnes en Schroeder (1967) [4]

Charnes en Schroeder behandelen een toepassing, gelijksoortig aan de militaire toepassing uit de vorige sektie. Hun model is gelijk aan model 2^b. De toestanden in hun model komen overeen met de mogelijke informatiestaten van de achtervolger. De aktieruimten zijn de taktische plannen.

Ze eisen, dat het model stoppend is in die zin, dat onder ieder paar stationaire strategieën vanuit iedere toestand er een positieve kans is, dat het spel stopt in één stap.

Charnes en Schroeder analyseren dit model voor twee verschillende doelstellingen. Ten eerste het minimaliseren door de zoeker (maximaliseren door de vluchter) van de verwachte duur van het spel. Ten tweede het maximaliseren door de zoeker (minimaliseren van de vluchter) van de verwachte kans op pakken.

Deze tweede doelstelling beschouwen ze tevens onder de conditie, dat de zoeker zodanig speelt, dat de maximale verwachte duur van het spel kleiner is dan een bepaalde waarde.

Daar hun model per stap positieve stopkansen heeft, komt dit neer op de lineaire nevenvoorwaarden, die onder model 2^b zijn gegeven voor het geval T=1 (zie (2.12)).

Vervolgens beschouwen ze nog het geval van perfecte informatie voor de zoeker, waarmee ze bedoelen, dat de vluchter volgens een bepaalde, aan de zoeker bekende, strategie wegvlucht. Aldus verkrijgen ze model 1^b en ze komen dan ook tot de conclusie, dat de zoeker nu zuivere optimale stationaire strategieën bezit.

Ten slotte wijden ze nog enkele beschouwingen aan het eindig-staps spel (modellen 1^a en 2^a).

c. Klein (1968) [9].

Het model, dat Klein beschouwt, is een uitbreiding van model 1^a. De verstopper verstopt zich na iedere periode opnieuw met een voor de zoeker bekende kansverdeling. Aan de oplosmethode, zoals onder model 1^a in sektie 2.3 aangegeven, is te zien, dat deze uitbreiding neerkomt op het invoeren van niet-stationaire direkte uitbetalingen en niet-stationaire overgangskansen. De methode kan rechtstreeks aan-

gepast worden.
De beperking op het aantal stappen in het model van Klein wordt geforceerd door een beperking op het budget.
Klein analyseert dit model door het te beschouwen als een spel met oneindige horizon, namelijk zo gauw het spel stopt, laat hij het weer in de begintoestand starten. Een markov-strategie voor het eindigstaps spel is dan een stationaire strategie voor het spel met oneindige horizon met een bijbehorende Markov-keten met één kernfuik.
Met behulp van de stationaire kansverdelingen kan Klein alle relevante grootheden, zoals verwachte kosten, verwachte duur van het spel, verwachte vindkans, enz. bepalen.
Voor doelstellingen van het type 1,2 of 3 (zie sektie 2.2) adviseert hij een dynamisch programmeringsoplossingstechniek.
Voor doelstellingen van het type 4, dus met constraints adviseert hij de lineaire programmeringsmethode, zoals o.a. uitgewerkt in Derman [5].

d. Sweat (1970) [18]

Het model van Sweat is analoog aan model 1^b. Zijn verdisconteringsfactor is afhankelijk van de toestand en met name van de doos, waarin in de vorige periode gezocht is. De verdisconteringsfactor wordt aldus gezien als een soort overlevingskans van de zoeker.
De doelstelling is het maximaliseren van de vindkans.
Bestaan en uniciteit van de oplossing volgen op dezelfde manier als onde model 1^b.
Vervolgens laat Sweat zien, dat er een optimale strategie bestaat van een speciaal type, namelijk in toestand $\overline{q}$, zuivere aktie i kiezen, waarvoor $\frac{p_i q_i}{1-\beta_i}$ maximaal is ($\overline{q}$ is de conditionele verstopkansverdeling, $\overline{p}$ is de conditionele vindkans en $\overline{\beta}$ is de verdisconteringsfactor).
Ten slotte leidt Sweat nog enkele resultaten af voor de situatie, waarin een aantal (of alle) van de β_i's naar 1 gaan.

e. Sakaguchi (1972) [13]

Sakaguchi behandelt twee modellen, die beiden soortgelijk zijn

aan model 2^b. In beiden is de doelstelling het optimaliseren van de kosten en in beiden worden toekomstige kosten verdisconteerd.
Het eerste model kent maar twee toestanden, gepakt zijn of nog niet gepakt zijn. Sakaguchi geeft analytische uitdrukkingen voor de waarde van het spel en voor de optimale strategieën (volledig gemengd) voor de beide spelers. Deze formules hebben nauwe verwantschap met de formules, die Neuts geeft voor zijn model.
Het merkwaardige feit doet zich voor, dat de optimale strategie voor de zoeker onafhankelijk is van de kosten voor het zoeken in de diverse doosjes.
In het tweede model van Sakaguchi zijn zowel de verstopper als de zoeker 'luidruchtig' en 'niet-doof'. De toestandsruimte bestaat dientengevolge uit de paren (k_1,k_2) met k_1 de doos, waarin de verstopper zich de vorige periode verstopt heeft en k_2 de doos, waarin de zoeker gezocht heeft. Kosten, opbrengsten, vindkansen, enz. zijn afhankelijk van de doos.
Dit model is volkomen analoog aan model 2^b.
Voor een model met twee dozen worden analytische uitdrukkingen voor de oplossing gegeven.

HOOFDSTUK 3

Toepassing van de theorie der stochastische spelen op duopoly modellen

3.1. Inleiding

In dit hoofdstuk zullen we beschrijven hoe stochastische spelmodellen behulpzaam kunnen zijn bij het analyseren van een duopoly situatie. We beschouwen een situatie waarbij twee producenten hetzelfde produkt op de markt brengen. De tijdas denken we ons opgedeeld in een aantal (mogelijk oneindig veel) even grote perioden. Laat de toestand van het systeem bepaald zijn door de huidige voorraden, de huidige prijzen, de huidige vraag, het produktieniveau, de reclameuitgaven, de kwaliteit van het produkt, enz. Laat s de variabele zijn waarin al deze grootheden voor de beide duopolisten in tot uitdrukking komen.

Laat S de verzameling van waarden zijn die s kan aannemen. Laat in toestand s $A_n(s)$ de verzameling van mogelijke akties zijn die duopolist n kan nemen. Deze kunnen bestaan uit het aanpassen van de prijs, het veranderen van het produktieniveau, enz.

De beide producenten kiezen aan het einde van een periode gelijktijdig en onafhankelijk een aktie. Het gevolg van deze beide akties is een nieuwe situatie, waaruit een vraag v_1, naar produkten van de eerste producent en een vraag v_2 naar produkten van de tweede producent voortvloeit. Echter welke v_1 en v_2 optreden is niet met zekerheid bekend; wel is bij gegeven akties aan beide spelers bekend een kansverdeling op de verzameling van alle mogelijke paren (v_1,v_2).

Laat V de verzameling van alle mogelijke paren (v_1,v_2) zijn. We veronderstellen dat V uit een eindig aantal elementen bestaat. Voor gegeven (v_1,v_2) kan met behulp van de gekozen akties en de huidige toestand, de toestand aan het einde van de komende periode berekend worden. De gekozen akties en de huidige toestand bepalen dus een kansverdeling op de toestandsverzameling S voor de toestand die aan het einde van de komende periode kan optreden.

Als we zowel S, $A_1(s)$, $A_2(s)$ als V eindig veronderstellen, dan hebben we niets anders geformuleerd dan een eindig twee-persoons stochastisch spel. Dit spel is niet-nulsom. De gekozen akties geven een direkte opbrengst voor

een producent, die gelijk is aan de winst in de komende periode, welke gelijk is aan de inkomsten door de verkoop van zijn produkten verminderd met de kosten die hij in de betreffende periode maakt.
Daar de vraag stochastisch van de akties afhangt, hangen ook de inkomsten stochastisch van de akties af, zodat we van een verwachte winst moeten spreken.
We beschouwen het verdisconteerde model; er is een getal β, $0 \leq \beta < 1$ gespecificeerd; een bedrag $r(t)$ dat betrekking heeft op de t^e periode wordt geacht β^{t-1} $r(t)$ waard te zijn aan het begin van de eerste periode. Stelling 1.3.2 uit hoofdstuk 1 laat zien dat boven beschreven spel minstens één evenwichtspunt bezit van stationaire strategieën. In het algemeen zijn er meerdere evenwichtspunten; tevens is het moeilijk om dezen te bepalen. In de volgende sektie zal aan deze moeilijkheden nadere aandacht worden geschonken.

OPMERKING. We kunnen dit model als volgt uitbreiden tot een model, waarbij de duopolisten niet gelijktijdig hun aktie hoeven te kiezen. Beschouw een situatie waarbij beide producenten één keer per dag (bijvoorbeeld 's avonds) een aktie mogen maar niet hoeven kiezen. Zodra aan het einde van een dag minstens één van de beide producenten een aktie kiest, verandert de toestand van het systeem en vangt een nieuwe periode aan. We nemen in eerste instantie aan dat T het aantal dagen is, waarbinnen de duopolisten mintens één keer een aktie moeten kiezen. We kunnen nu dit model zien als een stochastisch spel met niet-constante perioden. De aktieruimte voor speler n in toestand s is nu $\{1,2,...,T\} \times A_n(s)$, d.w.z. speler n kiest zowel een dag waarop hij zal beslissen als een aktie die hij dan zal nemen. Als de dagen waarop de producenten hun aktie wensen uit te voeren niet gelijk zijn, wordt alleen die aktie uitgevoerd van de producent die het eerst iets wilde ondernemen, waardoor op de betreffende dag een toestandsverandering plaatsvindt. Daarna plannen beiden weer een dag waarop ze vervolgens iets zouden willen ondernemen en een bijbehorende aktie, enz. Door de niet-constante periodeduur moet de verdisconteringsfactor steeds aangepast worden. Een uitwerking van dit model is te vinden in VRIEZE [20]. Als er geen periode T is waarbinnen de producenten minstens één keer een aktie moeten kiezen, dan hebben we te maken met overaftelbare aktieruimten. Ook hiervoor zijn resultaten binnen de theorie der stochastische spelen bekend, zie bijvoorbeeld FEDERGRÜN [6] of VRIEZE [19].

3.2. Myopic evenwichtspunten in duopoly modellen

In deze sektie komen zogenaamde myopic evenwichtspunten aan de orde. Hiermee worden evenwichtsstrategieën bedoeld, waarbij de beide spelers steeds slechts één periode vooruit hoeven te kijken. De resultaten in deze sektie zijn grotendeels gebaseerd op werk van SOBEL [15], [16] en [17]. Het model waar we vanuit gaan is gelijk aan het model van sektie 3.1, dus met S, $A_1(s)$, $A_2(s)$ en V allen eindige verzamelingen. De te gebruiken variabelen hebben de volgende betekenis:

Een index n heeft steeds betrekking op speler n, n = 1,2.
Een index t heeft steeds betrekking op periode t, t = 1,2,.. .
p_{nt}: prijs voor produkten
v_{nt}: vraag naar produkten
z_{nt}: voorraad aan het begin van de periode
w_{nt}: produktiegrootte
y_{nt}: $y_{nt} = z_{nt} + w_{nt}$, dus totale hoeveelheid ter beschikking staande produkten in de periode
$a_n(s)$: aktie in toestand s
$g_n(a_n(s),s)$: kosten als in toestand s aktie $a_n(s)$ genomen wordt
β_n: verdisconteringsfactor.

Als speler n in periode t besluit w_{nt} te produceren, dan veronderstellen we, dat deze produktie dadelijk aan het begin van de periode plaatsvindt. We gaan uit van een model waarbij negatieve voorraden mogen optreden. Een vraag waaraan in een bepaalde periode niet voldaan kan worden, wordt in de volgende periode verzorgd.

AANNAME 1. De kansverdeling op de vraag in de komende periode bij akties $a_1(s)$ en $a_2(s)$ in toestand s is uitsluitend afhankelijk van de voorraden dadelijk na het begin van de periode (y_{1t} en y_{2t}) en van de prijzen (p_{1t} en p_{2t}).

We kunnen de kosten $g_n(a_n(s),s)$ opdelen in drie komponenten: c_{n1}, c_{n2} en

c_{n3}. Hierbij is c_{n1} het deel dat betrekking heeft op het verschil tussen de voorraad dadelijk na het begin van de periode (y_{nt}) en de vraag v_{nt} in de periode.
c_{n2} zijn de kosten voor het produceren; c_{n3} zijn de overige kosten.

AANNAME 2. $c_{n2} = a_n \cdot w_{nt}$ dus lineair in de produktiegrootte $c_{n3} = b_n\, w_{nt} + f_n(y_{nt}, v_{nt})$, dus uitsluitend afhankelijk van w_{nt}, y_{nt} en v_{nt} en wel lineair in w_{nt}; op de functie $f_n(\cdot,\cdot)$ liggen geen restricties.

Laat $\bar{y}_t = (y_{1t}, y_{2t})$ en $\bar{p}_t = (p_{1t}, p_{2t})$. Laat $p_n(\cdot \mid \bar{y}_t, \bar{p}_t)$ de kansverdeling op de vraag zijn voor speler n bij $\bar{y}_t$ en $\bar{p}_t$. Als in toestand s_t in periode t de spelers respektievelijk $a_1(s)$ en $a_2(s)$ kiezen dan volgt uit het bovenstaande dat we de verwachte winst in deze periode voor speler n (notatie $V_{nt}(s_t, \bar{a}(s))$ met $\bar{a}(s) = (a_1(s), a_2(s))$) kunnen schrijven als:

$$V_{nt}(s_t, \bar{a}(s_t)) = \sum_{v_n} p_n(v_{nt} \mid \bar{y}_t, \bar{p}_t)\{p_{nt} \cdot v_{nt} - g_n(a_n(s_t), s_t)\}$$

$$= \sum_{v_n} p_n(v_{nt} \mid \bar{y}_t, \bar{p}_t)\{p_{nt} \cdot v_{nt} - c_{n1}(y_{nt}, v_{nt}) - a_n w_{nt} - b_n w_{nt} - f_n(y_{nt}, v_{nt})\}$$

$$= \sum_{v_n} p_n(v_{nt} \mid \bar{y}_t, \bar{p}_t)\{p_{nt} \cdot v_{nt} - d_n \cdot w_{nt} - h_n(y_{nt}, v_{nt})\},$$

waarbij $d_n = a_n + b_n$ en $h_n(\cdot,\cdot) = c_{n1}(\cdot,\cdot) + f_n(\cdot,\cdot)$. Daar $y_{nt} = z_{nt} + w_{nt}$ blijkt dat deze uitdrukking uitsluitend afhankelijk is van $\bar{y}_t$ en $\bar{p}_t$ en van de toestand s_t via $\bar{z}_t$. Voor speler n kunnen we in toestand s dus als aktieruimte alle paren (y_n, p_n), die mogelijk zijn, nemen. De enige relevante grootheden uit de toestandsvariabele s zijn z_1 en z_2, zodat in het vervolg een toestand aangeduid wordt met een paar (z_1, z_2). Een stationaire strategie voor speler n is dan het aan ieder paar (z_1, z_2) toevoegen van een paar (y_n, p_n) met $y_n \in \{K(z_1), K(z_1)+1, \ldots, Y_n\}$, $p_n \in \{0, 1, \ldots, P_n\}$ waarbij $K(z_n) = \max\{0, z_n\}$, Y_n het maximale niveau waarop de voorraad gebracht kan worden en P_i de maximale prijs is.
De totale verwachte verdisconteerde opbrengst voor speler n bij een starttoestand (z_{10}, z_{20}) onder een paar stationaire strategieën is dan te schrijven als:

$$W_n((z_{10},z_{20}),\pi_1,\pi_2) =$$

$$E_{\pi_1\pi_2}\Big[\sum_{t=1}^{\infty} \beta_n^{t-1} \sum_{v_n} p(v_{nt}|\bar{y}_t,\bar{p}_t)\{p_{nt}\cdot v_{nt} - d_n w_{nt} - h_n(y_{nt},v_{nt})\}\Big] =$$

$$E_{\pi_1\pi_2}\Big[\sum_{t=2}^{\infty} \beta_n^{t-1} \sum_{v_n} p(v_{nt}|\bar{y}_t,\bar{p}_t)\{p_{nt}\cdot v_{nt} - d_n(y_{nt}-(y_{nt-1}-v_{nt-1})) - h_n(y_{nt},v_{nt})\}\Big]$$

$$+ \sum_{v_n} p(v_{n1}|\bar{y}_1,\bar{p}_1)\{p_{n1}\cdot v_{n1} - d_n y_{n1} + d_n z_{n0} - h_n(y_{n1},v_{n1})\} =$$

$$E_{\pi_1\pi_2}\Big[\sum_{t=1}^{\infty} \beta_n^{t-1} \sum_{v_n} p(v_{nt}|\bar{y}_t,\bar{p}_t)\{p_{nt}\cdot v_{nt}-d_n(1-\beta_n)y_{nt} - d_n\beta_n v_{nt} - h_n(y_{nt},v_{nt})\}\Big]$$

$$- d_n z_{n0}.$$

Hier moeten $\bar{y}_t$, $\bar{p}_t$ en w_{nt} opgevat worden als stochastische variabelen, waarvan de kansverdeling bepaald wordt door de stationaire strategieën π_1 en π_2 van de beide spelers.
De term $d_n z_{n0}$ is constant voor begintoestand (z_{10},z_{20}) en hoeft dus voor het optimaliseren van $W_n((z_{10},z_{20}),\pi_1,\pi_2)$ niet in aanmerking genomen worden. Wat overblijft is een uitdrukking waarbij de t^e term uitsluitend nog afhankelijk is van de gekozen akties op tijdstip t en niet meer van de toestand (wel indirekt daar de toestand de aktieruimte bepaald).

Aan iedere toestand (z_1,z_2) voegen we nu een bimatrixspel $(A_1(z_1,z_2),A_2(z_1,z_2))$ toe, waarbij de aktieruimte voor speler n gelijk is aan de zuivere aktieruimte voor speler n in toestand (z_1,z_2) en de uitbetaling aan speler n bij zuivere akties (y_1,p_1), (y_2,p_2) gelijk is aan

$$\alpha_n(y_1,p_1,y_2,p_2) =$$

$$\sum_{v_n} p(v_n|\bar{y},\bar{p})\{p_n\cdot v_n - d_n(1-\beta_n)y_n - d_n\beta_n v_n - h_n(y_n,v_n)\}.$$

Laten Z_1 en Z_2 de kleinste voorraden zijn die kunnen optreden (Z_1 en Z_2

kunnen negatief zijn).

STELLING 3.2.1. *Indien het bimatrixspel* $(A_1(Z_1,Z_2),A_2(Z_1,Z_2))$ *een evenwichtspunt van zuivere strategieën bezit, zeg* $(y_1^*,p_1^*),(y_2^*,p_2^*)$*, dan vormen de stationaire strategieën* π_1^* en π_2^* *een evenwichtspunt voor de begintoestanden* (z_1,z_2)*, met* $z_1 \le y_1^*$ en $z_2 \le y_2^*$*, als geldt:* $\pi_n^*(z_1,z_2) = (y_n^*,p_n^*)$ *voor* $z_n \le y_n^*$ *en* $\pi_n^*(z_1,z_2)$ *is willekeurig in de toestanden met* $z_n > y_n^*$, $n = 1,2$.

BEWIJS. In de eerste plaats kunnen we opmerken dat $(y_1^*,p_1^*),(y_2^*,p_2^*)$ ook evenwichtspunt is in de bimatrixspelen $(A_1(z_1,z_2),A_2(z_1,z_2))$ met $z_1 \le y_1^*$ en $z_2 \le y_2^*$.
Voor een begintoestand (z_{10},z_{20}) met $z_{n0} \le y_n^*$ is de strategie π_n^* zodanig dat voor iedere t $z_{nt} \le y_n^*$.
Voor een willekeurige stationaire strategie voor speler 1 geldt dan:

$$
\begin{aligned}
W_1((z_{10},z_{20}),\pi_1,\pi_2^*) &= E_{\pi_1\pi_2^*}\left[\sum_{t=1}^{\infty} \beta_1^t\cdot\alpha_1(y_{1t},p_{1t},y_{2t}^*,p_{2t}^*)\right] + d_1z_{10} \\
&\le E_{\pi_1\pi_2^*}\left[\sum_{t=1}^{\infty} \beta_1^t\cdot\alpha_1(y_{1t}^*,p_{1t}^*,y_{2t}^*,p_{2t}^*)\right] + d_1z_{10} \\
&= E_{\pi_1^*\pi_2^*}\left[\sum_{t=1}^{\infty} \beta_1^t\cdot\alpha_1(y_{1t}^*,p_{1t}^*,y_{2t}^*,p_{2t}^*)\right] + d_1z_{10} \\
&= W_1((z_{10},z_{20}),\pi_1^*,\pi_2^*).
\end{aligned}
$$

Evenzo geldt $W_2((z_{10},z_{20}),\pi_1^*,\pi_2) \le W_2((z_{10},z_{20}),\pi_1^*,\pi_2^*)$, $\forall\pi_2$, waaruit blijkt dat π_1^*,π_2^* een evenwichtspunt vormen. □

Noteer een gemengde aktie in het bimatrixspel $(A_1(z_1,z_2),A_2(z_1,z_2))$ voor speler n als $\pi_n(\cdot,\cdot)$, dus $\pi_n(y_n,p_n) \ge 0$ en $\sum_{(y_n,p_n)} \pi_n(y_n,p_n) = 1$. Laat bij iedere $\pi_n(\cdot,\cdot)$ $y_n(\pi_n) = \min_{y_n}\{y_n \mid \exists p_n \text{ met } \pi_n(y_n,p_n) > 0\}$.

STELLING 3.2.2. *Indien in het bimatrix spel* $(A_1(Z_1,Z_2),A_2(Z_1,Z_2))$ *het paar* $\pi_1^*(\cdot,\cdot)$ *en* $\pi_2^*(\cdot,\cdot)$ *een evenwichtspunt vormt, dan vormen de strategieën* π_1^* *en* π_2^* *een evenwichtspunt voor alle begintoestanden* (z_1,z_2) *met* $z_1 \le y_1(\pi_1^*)$ *en* $z_2 \le y_2(\pi_2^*)$*, als geldt:*

a) *Voor iedere toestand* (z_1,z_2) *met* $z_n \leq y_n(\pi_n^*)$ *stemt* $\pi_n^*(z_1,z_2)$ *overeen met* $\pi_n^*(\cdot,\cdot)$.

b) *Voor alle paren* (y_1,y_2) *en alle paren* (p_1,p_2), *zodanig dat* $\pi_1^*(y_1,p_1) > 0$ *en* $\pi_2^*(y_2,p_2) > 0$ *geldt dat zowel*

$$\sum_{v_1 \geq y_1 - y_1(\pi_1^*)} p_1(v_1|\bar{y},\bar{p}) = 1 \quad als \quad \sum_{v_2 \geq y_2 - y_2(\pi_2^*)} p_2(v_2|\bar{y}.\bar{p}) = 1.$$

BEWIJS. Het bewijs is analoog aan het bewijs van Stelling 3.2.1. Voorwaarde b) zorgt ervoor dat voor een starttoestand (z_{10},z_{20}) met $z_{n0} \leq y_n(\pi_n^*)$ het systeem onder strategie π_n^* uitsluitend toestanden doorloopt met $z_{nt} \leq y_n(\pi_n^*)$. □

De evenwichtspunten uit de stellingen 3.2.1 en 3.2.2 worden myopic evenwichtspunten genoemd, omdat de spelers per stap steeds evenwichtspunten voor de komende periode spelen.

OPMERKING. De stellingen 3.2.1 en 3.2.2 blijven ook gelden als we het t-staps spel (al of niet verdisconteerd) beschouwen of het spel, waarbij we de gemiddelde opbrengst per periode als criterium nemen.

In [15] behandelt SOBEL bovenstaand spel voor het geval, waarbij de beide spelers zowel het nieuwe voorraadniveau y als de prijs p uit een interval mogen kiezen. Onder voldoende voorwaarden op de functies $d_n(\cdot,\cdot,\cdot,\cdot)$, $n = 1,2$ formuleert hij stellingen analoog aan de beide bovenstaande stellingen.

3.3. Unieke evenwichtspunten

In deze sektie wordt op het model zoals behandeld in de vorige sektie meer restrikties aangebracht, zodanig dat we een uniek evenwichtspunt verkrijgen.
De inhoud van deze sektie is gebaseerd op enige resultaten uit VRIEZE [20].
We richten onze aandacht eerst op bimatrixspelen.

DEFINITIE. Een matrix A wordt *concaaf in de kolommen (rijen)* genoemd, als geldt: $a_{ij}-a_{i-1j} > a_{i+1j}-a_{ij}$, $i = 2,\ldots,m-1$, $j = 1,\ldots,n$

$$(a_{ij} - a_{ij-1} > a_{ij+1} - a_{ij}, \quad j = 2,\ldots,n-1, \quad i = 1,\ldots,m).$$

DEFINITIE. Een matrix A wordt *één-toppig in de kolommen (rijen)* genoemd als er bij iedere kolom j (rij i) een rij i_0 (kolom j_0) is, z.d.d. $a_{i_0 j} > a_{ij}$, $\forall i \neq i_0$

$$(a_{ij_0} > a_{ij}, \quad \forall j \neq j_0).$$

In het vervolg noteren we voor een matrix A die één-toppig in de kolommen (rijen) is de rij (kolom) waarvoor voor kolom j (rij i) het maximum aangenomen wordt als i_j (j_i).

DEFINITIE. Een matrix A, die één-toppig in de kolommen (rijen) is, wordt *langzaam-top-dalend (langzaam-top-stijgend)* genoemd als geldt: óf $i_j = i_{j-1}$ óf $i_j = i_{j-1}-1$, $j = 2,\ldots,n$

$$(\text{óf } j_i = j_{i-1} \quad \text{óf } j_i = j_{i-1}+1, \quad i = 2,\ldots,m).$$

STELLING 3.3.1. *Een bimatrixspel* (A,B) *waarvoor de matrix* A *concaaf in de kolommen, ééntoppig in de kolommen en langzaam-top-dalend is en waarvoor de matrix* B *concaaf in de rijen, ééntoppig in de rijen en langzaam-top-stijgend is bezit een uniek evenwichtspunt, waarbij voor beide spelers de evenwichtsakties hoogstens gewicht leggen op twee zuivere akties.*

BEWIJS. Het bewijs is nogal technisch van aard en is identiek gelijk aan het bewijs van stelling 3.1 uit VRIEZE [20] en zal daarom niet gegeven worden. □

Onderstaande figuren geven inzicht in wat er aan de hand is bij bimatrix-spelen die voldoen aan de in de stelling genoemde eisen.

A B

In dit (4×4)-bimatrix-spel stellen, voor respektievelijk de matrices A en B, de gebroken lijnen de verbindingen voor tussen de toppen van respektievelijk de kolommen en de rijen. Met behulp van deze figuren is in te zien dat precies één van de volgende twee mogelijkheden optreedt:

a) Er bestaat een rij i_0 en een kolom j_0, z.d.d.

$$a_{i_0 j_0} > a_{i j_0}, \quad \forall i \neq i_0 \quad \text{en} \quad b_{i_0 j_0} > b_{i_0 j}, \quad \forall j \neq j_0.$$

b) Er bestaat een rij i_0 en een kolom j_0, z.d.d.

$$b_{i_0 j_0} > b_{i_0 j}, \quad \forall j \neq j_0 \qquad b_{i_0+1\, j_0+1} > b_{i_0+1\, j_0}, \quad \forall j \neq j_0+1$$

$$a_{i_0 j_0+1} > a_{i j_0+1}, \quad \forall i \neq i_0 \qquad a_{i_0+1\, j_0} > a_{i j_0}, \quad \forall i \neq i_0+1.$$

In de figuur treedt geval b) op.

In geval a) bestaat het evenwichtspunt bedoeld in de stelling uit de zuivere akties i_0 voor speler 1 en j_0 voor speler 2. In geval b) bestaat het evenwichtspunt uit gemengde akties, die alleen gewicht leggen op respektievelijk de rijen i_0 en i_0+1 en de kolommen j_0 en j_0+1. De gewichten op deze rijen en kolommen zijn gelijk aan de gewichten behorende bij het unieke gemengde evenwichtspunt van het (2×2)-bimatrixspel:

$$\begin{pmatrix} a_{i_0 j_0} & a_{i_0 j_0+1} \\ a_{i_0+1\, j_0} & a_{i_0+1\, j_0+1} \end{pmatrix} \quad \begin{pmatrix} b_{i_0 j_0} & b_{i_0 j_0+1} \\ b_{i_0+1\, j_0} & b_{i_0+1\, j_0+1} \end{pmatrix}$$

We gaan nu terug naar het stochastische spel zoals beschouwd in sektie 3.2.

We zeggen dat dit stochastisch spel de *geslaagde rangschikkingseigenschap* heeft als de zuivere akties (y_n, p_n) voor speler n bij toestand (Z_1, Z_2) zodanig te rangschikken zijn, dat de waarde van y_n niet dalend is en dat het bimatrixspel met elementen berekend met behulp van $\alpha_1(y_1,p_1,y_2,p_2)$ en $\alpha_2(y_1,p_1,y_2,p_2)$ volgens deze nieuwe rangschikking voldoet aan de eigenschappen vermeld in stelling 3.3.1.

<u>STELLING 3.3.2.</u> *Een stochastisch spel zoals geformuleerd in sektie* 3.2, *dat de geslaagde rangschikkingseigenschap heeft, bezit een uniek even-*

wichtspunt voor begintoestanden (z_{10}, z_{20}) *waarvoor* z_{10} *en* z_{20} *voldoende klein zijn en waarvoor de vraag in iedere periode voor beide producenten minstens 1 eenheid produkt is.*

BEWIJS. Het bewijs komt neer op het kombineren van de stellingen 3.2.2 en 3.3.1. □

De voorwaarde dat beide producenten in iedere periode een vraag van minstens 1 eenheid produkt moeten hebben, is alleen van toepassing als het evenwichtspunt niet een evenwichtspunt van zuivere strategieën is.
Dat een dusdanig stochastisch spel heel wel de geslaagde rangschikkingseigenschap kan hebben toont het volgende voorbeeld.

VOORBEELD. Beide producenten kunnen kiezen uit twee produktieniveaus, "laag" en "hoog" aangeduid met l en h. Beide producenten kunnen kiezen uit drie prijzen, "extra laag", "laag" en "hoog" aangeduid met e.l., l en h.
Een geslaagde rangschikking zou er als volgt uit kunnen zien:

	lh	ll	hh	hl	he.l
lh					
ll					X
hl		X	X	X	
hh	X				
he.l					

A

	lh	ll	hh	hl	he.l
lh			X		
ll				X	
hl				X	
hh				X	
he.l					X

B

De kruisjes geven de plaatsen aan, waar voor de matrices A en B respektievelijk de maxima in de kolommen en de maxima in de rijen optreden.
Een extra lage prijs kan opgevat worden als een stuntprijs, waarbij amper nog winst gemaakt wordt. Blijkbaar kan speler 2 zijn stuntaanbiedingen beter kwijt dan speler 1.
Ga na dat ook het concaaf zijn van de kolommen van A en de rijen van B alleszins denkbaar is.

het unieke evenwichtspunt in dit bimatrixspel is het paar ((h,l),(h,l)), dus beide spelers hebben een hoge produktie en een lage prijs.

LITERATUUR

[1] AUMANN, R.J., *Mixed and behaviour strategies in infinite extensive games, Advances in Game Theory*, Ann. Math. Studies no 52, Princeton Univ. Press, Princeton (1964) 627-650.

[2] BELLMAN, R., *Dynamic programming,* Princeton Univ. Press, Princeton (1957).

[3] BLACKWELL, D., *Discounted dynamic programming,* Ann. Math. Stat. 36 (1965) 226-235.

[4] CHARNESS, A.M. & R.C. SCHROEDER, *On some stochastic tactical anti-submarine games,* Naval Res. Log. Quaterly 14 (1967) 291-312.

[5] DERMAN, C., *Optimal replacement under Markovian deterioration with probability bounds on failure,* Management Science 9 (1963).

[6] FEDERGRÜN, A., *On N-person games with denumerable state space,* report BW 67/76, Mathematisch Centrum, Amsterdam, 1976.

[7] HOWARD, R., *Dynamic programming and Markov Processes,* MIT Press, Cambridge Massachusets.

[8] KAKUTANI , S., *A generalisation of Brouwer's fixed point theorem,* Duke Math. J. 7 (1941) 457-459.

[9] KLEIN, M., *Note on sequential search,* Naval Res. Log. Quaterly 15 (1968) 469-475.

[10] NEUMANN, J. VON, *Zur Theorie der Gesellschaftsspiele,* Mathematische Annalen 100 (1928) 295-320.

[11] NEUTS, M.F., *A multistage search game,* J. SIAM, Appl. Math. 11 (1963) 502-507.

[12] ROGERS, P.D., *Non-zero-sum stochastic games*, Report ORC 69-8, Operations Research Centre, Univ. of California, Berkely: Ph.D. dissertation (1969).

[13] SAKAGUCHI, M., *Two-sided search games*, J. of the Operations Research Society of Japan 16 (1972) 207-225.

[14] SHAPLEY, L.S., *Stochastic games,* Proc. Nat. Acad. Sc. 39 (1953) 1095-1100.

[15] SOBEL, M.J. & A.P. KIRMAN, *Dynamic oligopoly with inventories,* Econometrica 42 (1974) 279-287.

[16] SOBEL, M.J., *Myopic equilobrium points in sequential games*, submitted to J. of Economic Theory (1976).

[17] SOBEL, M.J., *Simply solved stochastic games*, unpublished manuscript (1977).

[18] SWEAT, C.W., *Sequential search with discounted income, the discount a function of the cell searched,* Ann. Math. Stat. 41 (1970), 1446-1455.

[19] VRIEZE, O.J., *The stochastic non-cooperative countable person game with countable state space and compact action spaces under the discounted payoff criterium*, Report BW 66/76, Mathematisch Centrum, Amsterdam (1976).

[20] VRIEZE, O.J., *Duopoly models, stochastic games and bimatrix games*, Report BW 76/77, Mathematisch Centrum, Amsterdam (1977).

[21] WAL, J. VAN DER, *The method of successive approximation for the discounted Markov-game*, Technological University Eindhoven (Department of Mathematics, Memorandum COSOR 75-02 (1975).

[22] WAL, J. VAN DER, *Successive approximation and discounted Markov games,* University of Technology Enschede (Department of Applied Mathematics), memorandum nr. 119 (1976).