Hoe te gokken als het moet

Tim van Wingerden

20 januari 2006

Inhoudsopgave

1 Inleiding en motivatie

Voor het kiezen van dit onderwerp heb ik vooral gekeken naar welke richting me het meest interessant leek. Aangezien ik al een aantal Besliskundevakken heb gevolgd lag het voor de hand in die richting mijn bachelorscriptie te gaan schrijven. Na overleg met dr. L.C.M Kallenberg heb ik besloten om als uiteindelijk onderwerp zogenaamde 'casino-modellen' te gaan bestuderen. De scriptie zelf heb ik onder leiding van dr. Spieksma geschreven. Ik heb bij het schrijven van de scriptie vooral gebruik gemaakt van de lecture notes [3] en 'Linear Programming and future Markovian control problems' [2] van dr. L.C.M. Kallenberg en verder van het boek 'How to gamble if you must' [1] van Dubins en Savage. Het uiteindelijke doel van het onderzoek bestaat hoofdzakelijk uit het bestuderen van Red & Black modellen, het vinden van optimale strategieën en te bewijzen dat deze inderdaad optimaal zijn. Concreet komt dit neer op de volgende punten:

1) Het gokmodel schrijven als een Markov Beslissingsketen met eindige toestands- en actieruimte en met totale opbrengstencriterium;

2) bewijzen dat het totale opbrengstenmodel in feite een transiënt model is;

3) aantonen dat de optimaliteitsvergelijking een unieke oplossing heeft en dat het successieve approximatiealgoritme convergeert;

4) analytische oplossingen vinden voor het gokmodel voor verschillende winstkansen;

5) berekeningen van waardevectoren uitvoeren met Matlab.

2 Model en notatie

De term casinomodel uit de inleiding doet al vermoeden dat het gaat om een persoon die naar een casino gaat en daar gaat gokken. Deze speler heeft een startkapitaal bij zich en gaat vervolgens het spel Red & Black spelen. Daarbij kan hij een bedrag inzetten, uiteraard kleiner dan het kapitaal dat hij op dat moment heeft. Het doel van de speler is om een vooraf bepaald streefbedrag te halen, zeg N. Om onnodig risico te vermijden, betekent dit dat hij nooit meer in zal zetten dan precies genoeg om het streefbedrag te halen. Hij stopt met spelen als hij het streefbedrag N heeft bereikt en ook als hij geen kapitaal meer heeft. Dit heeft tot gevolg dat de speler in feite een gokstrategie wil gebruiken die de kans dat hij N haalt maximaliseert.

Dit laat zich dan modelleren als een Markov beslissingsketen (MDP) met eindige toestandsruimte en actieruimte: $\mathbf{E} = \{0, 1, 2, .., N\}$ is de toestandsruimte. Toestand i staat voor kapitaal i en voor elk van deze i hoort een corresponderende actieverzameling $\mathbf{A}(i)$ van toegestane acties.

- Inzet 0 betekent dat het spel is gestopt: $\mathbf{A}(0) = 0$. De speler is blut en het spel is afgelopen
- Voor $i \neq 0, N$ is $\mathbf{A}(i) = \{1, 2, .., \min(i, N - i)\}$. De speler kan niet meer inzetten dan zijn kapitaal i. Hij wil ook niet meer inzetten dan precies genoeg om N te halen, dat is $N - i$.
- $\mathbf{A}(N) = 0$. De speler heeft zijn streefbedrag gehaald en stopt met spelen.

Zodra de speler zijn inzet heeft gekozen, wordt de volgende toestand bepaald door de kans op winst dan wel verlies: p_{iaj} de kans om van toestand i, bij keuze van inzet a, naar toestand j te gaan, dus in dit geval:

$$p_{iaj} = \begin{cases} p, & \text{als } j = i + a, i \neq 0, N \\ q = 1 - p, & \text{als } j = i - a, i \neq 0, N, \\ 0, & \text{voor } i = 0, j \in \mathbf{E}. \end{cases}$$

Verder zijn er opbrengsten r_{ia} wanneer in toestand i actie a is gekozen. In dit model, waar de speler de kans om N te bereiken wil maximaliseren, is de meest voor de hand liggende keuze:

$$r_{ia} = \begin{cases} 1, & \text{als } i = N, a = 0 \\ 0, & \text{anders.} \end{cases}$$

Laat $\boldsymbol{H}$ de verzameling met mogelijke toestanden en genomen acties van het systeem tot tijdstip t, dus

$$\begin{aligned} \boldsymbol{H}_t &= \{(i_1, a_1, i_2, a_2, .., i_{t-1}, a_{t-1}, i_t) |\, i_k \in \mathbf{E}, a_k \in \mathbf{A}(i_k), \\ &\quad k = 1, \ldots, t-1;\ i_t \in \mathbf{E}\}. \end{aligned}$$

Bijvoorbeeld $\boldsymbol{H}_2 = \{5, 1, 4\}$ betekent dat de speler begon met een kapitaal 5, toen $a = 1$ inzette, en verloor, zodat zijn kapitaal 4 werd.

Een *beslisregel* π^t op tijdstip t, is een niet-negatieve functie op $\boldsymbol{H}_t \times \mathbf{A}$, zodanig dat voor elke $h_t = (i_1, a_1, .., i_t) \in \boldsymbol{H}_t$ geldt dat $\pi^t_{h_t}(a_t) \geq 0$ voor $a_t \in \mathbf{A}(i_t)$, en $\sum_{a_t} \pi^t_{h_t}(a_t) = 1$. Een strategie $\mathbf{R}$ is een rij beslisregels.

Een strategie heet geheugenloos als π^t niet afhangt van (h_{t-1}, a_{t-1}), dus alleen van de laatste toestand i_t op tijdstip t. Dan kunnen we $\pi^t_{i_t}(a_t)$ schrijven in plaats van $\pi^t_{h_t}(a_t)$. Voorts heet een strategie *stationair*, als deze tijdsonafhankelijk is. Dat wil zeggen, zij hangt alleen af van de de huidige toestand: $\pi^t_{i_t}(a) = \pi^k_{i_k}(a)$ wanneer $i_t = i_k$. Tenslotte heet een beslisregel *deterministisch* als $\pi^t_{h_t, a_t} \in \{0, 1\}$, voor $i_t \in \mathbf{E}$ en $a_t \in \mathbf{A}(i_t)$. Zo'n beslisregel wordt volledig bepaald door de actie die in een toestand met kans 1 gekozen wordt. Je kunt een deterministische beslisregel dan ook als functie $f_t : \mathbf{E} \to \mathbf{A}$ schrijven. Een strategie die uitsluitend bestaat uit deterministische beslisregels heet een deterministische strategie. En een strategie die zowel deterministisch als stationair is, is dan ook te schrijven als een functie $f : \mathbf{E} \to \mathbf{A}$. In deze scriptie beperken we ons tot de klasse van de geheugenloze strategieën. En verder schrijf ik r_f voor de directe opbrengst onder beslisregel f en $P(f)$ voor de corresponderende overgangsmatrix.

3 Totale opbrengstenmodel

Een veel gebruikte manier om een bedrag uit het verleden te waarderen is door het gebruik van een rentepercentage r. Dan heeft een nu geïnvesteerd bedrag B op tijdstip t de waarde $(1+r)^t B$. Om een bedrag B op tijdstip t te hebben, is het dus nodig om $(1+r)^{-t}B$ te investeren. Laat nu $\alpha = (1+r)^{-1}$, dan is α een getal tussen 0 en 1 en α wordt de *verdisconteringsfactor* genoemd.

Verdiscontering stelt je dus in staat om opbrengsten over een oneindige tijdshorizon bij elkaar op te tellen tot een eindig getal. Voor $\alpha \in [0,1)$ definiëren we de (totale) verwachte verdisconteerde opbrengsten, gegeven begintoestand i en strategie $\mathbf{R}$ als volgt:

$$v_i^\alpha(\mathbf{R}) = \sum_{t=1}^{\infty} \sum_{i,a} \alpha^{t-1} \mathsf{P}_i^{\mathbf{R}}[X_t = j, Y_t = a] r_{ja},$$

waarbij X_t en Y_t stochastische variabelen zijn die respectievelijk de toestand en de gekozen actie op tijdstip t aangeven. Verder gebruiken we $\mathsf{P}_i^{\mathbf{R}}$ voor de kansoperator gegeven dat $x = i$ en strategie $\mathbf{R}$ wordt gebruikt.

Als we niet verdisconteren, d.w.z. $\alpha = 1$, dan krijg je een zogenaamd *totale opbrengstenmodel* (TOM). Ik neem voor voor de berekeningen ook aan dat de opbrengsten $r_{ia} \geq 0$. Het nadeel van een TOM-model is dat de totale opbrengsten onder een gegeven strategie niet altijd eindig hoeven te zijn en zelfs niet hoeven te bestaan. Laten we even aannemen dat deze eindig zijn voor elke strategie en begintoestand. Dan zijn de totale verwachte opbrengsten $v_i(\mathbf{R})$ onder strategie $\mathbf{R}$ en begintoestand i gedefinieerd als:

$$v_i(\mathbf{R}) = \mathsf{E}_i^{\mathbf{R}}(\sum_{t=1}^{\infty} r_{X_t Y_t}) = \sum_{t=1}^{\infty} \sum_{j,a} \mathsf{P}_i^{\mathbf{R}}[X_t = j, Y_t = a] r_{ja}.$$

Daarbij staat $\mathsf{E}_i^{\mathbf{R}}$ voor de verwachtingsoperator geassocieerd met $\mathsf{P}_i^{\mathbf{R}}$. Laat verder $v_i = \sup_{\mathbf{R}} v_i(\mathbf{R})$, dan heet een strategie $\mathbf{R}^*$ optimaal als geldt dat $v(\mathbf{R}^*) = v$ mits dit supremum bestaat.

4 Transiënt model

Bij ons gokmodel geldt voor de toestanden 0 en N dat $\sum_j p_{iaj} < 1$. Er verdwijnt dus eigenlijk kansmassa in die toestanden. Dit gebeurt typisch bij transiënte Markov beslissingsketens. Ons gokmodel is ook te schrijven als zo'n transiënte keten en deze is als volgt gedefinieerd. Een strategie $\mathbf{R}$ wordt transiënt genoemd als $\sum_{t,j} \mathsf{P}_i^{\mathbf{R}}(X_t = j) < \infty \, \forall i, j \in \mathbf{E}$. Het bewijs dat dit voor ons model ook geldt staat in het volgende lemma.

Lemma 4.1 *Voor het gokmodel geldt:*

$$\sum_{t,j} \mathsf{P}_i^{\mathbf{R}}(X_t = j) < \infty.$$

Bewijs.

$$\sum_{j=0}^{N} \mathsf{P}_i^{\mathbf{R}}(X_N = j) \leq 1 - \mathsf{P}_i^{\mathbf{R}}(\exists t < N \,|\, X_t \in \{0, N\}).$$

De laatste kans in de bovenstaande uitdrukking is de kans dat je voor tijdstip N het systeem "verlaten" hebt. We geven een onderafschatting voor deze kans:

$$\begin{aligned} \mathsf{P}_i^{R}(\exists t < N \,|\, X_t \in \{0, N\}) &\geq \mathsf{P}_i^{R}(\text{gokker verliest òf wint alleen maar}) \\ &= p^{N-i} + (1-p)^i \geq p^N + (1-p)^N. \end{aligned}$$

In de eerste ongelijkheid gebruik je, dat je in hooguit $N-1$ gokken het systeem vanuit toestand i altijd verlaat door ofwel altijd te winnen ofwel altijd te verliezen. In de derde maak je de macht van een getal kleiner dan 1 hoger, dus het totaal lager. Neem nu: $\beta = p^N + (1-p)^N$, dan volgt dat:

$$\mathsf{P}_i^{\mathbf{R}}(X_N = j) \leq 1 - \beta, \quad \forall \mathbf{R}, \forall i \in \mathbf{E}.$$

Dan volgt iteratief:

$$\begin{aligned} \sum_j P_i^R(X_{SN} = j) &\leq \sum_k \sum_j \mathsf{P}_i^{\mathbf{R}}(X_{SN} = j | X_{(S-1)N} = k) \cdot \mathsf{P}_i^{\mathbf{R}}(X_{(S-1)N} = k) \\ &\leq \sum_k (1-\beta) \cdot \mathsf{P}_i^{\mathbf{R}}(X_{(S-1)N} = k)) \\ &\leq \ldots \leq (1-\beta)^S. \end{aligned}$$

Voor $SN \leq t < (S+1)N$ geldt dan:

$$\begin{aligned} \sum_j \mathsf{P}_i^{\mathbf{R}}(X_t = j) &= \sum_{k=0}^{N} \mathsf{P}_i^{\mathbf{R}}(X_{SN} = k) \cdot \sum_j \mathsf{P}_k^{\mathbf{R}}(X_t = j | X_{SN} = k) \\ &\leq \sum_{k=0}^{N} \mathsf{P}_i^{\mathbf{R}}(X_{SN} = k) \leq (1-\beta)^S. \end{aligned}$$

De ongelijkheid volgt omdat de som van kansen kleiner of gelijk aan 1 is, en de laatste uit bovenstaande afschatting. Dit combinerend geeft weer dat:

$$
\begin{aligned}
\sum_t \sum_{j=0}^{N} \mathsf{P}_i^{\mathbf{R}}(X_t = j) &\leq \sum_{S=0}^{\infty} \sum_{t=SN}^{(S+1)N-1} \sum_{j=0}^{N} \mathsf{P}_i^{\mathbf{R}}(X_t = j) \\
&\leq \sum_{S=0}^{\infty} N(1-\beta)^S < N/\beta.
\end{aligned}
$$

□

We weten nu dus dat iedere strategie transiënt is en dus geldt dat dit model ook een transiënte Markov beslissingsketen is. Nu we dit bereikt hebben, kunnen we het volgende hulplemma afleiden. Dit is iets sterker dan Lemma 4.1, en veralgemeniseert een resultaat uit [2].

Lemma 4.2 *Voor een transiënte* MBK *met eindige toestands- en actieruimten geldt het volgende. Laat*

$$
\begin{cases}
y_i^1 &= 1, & i \in \mathbf{E} \\
y_i^{t+1} &= \max_{a \in \mathbf{A}(i)} \sum_j p_{iaj} y_j^t, & i \in \mathbf{E}, t = 1, \ldots.
\end{cases}
$$

Laat f^t de beslisregel zijn, die maximaliserende acties in de $(t+1)$-ste iteratie kiest, d.w.z.

$$
\sum_j p_{if^t(i)j} y_j^t = \max_a \sum_j p_{iaj} y_j^t, \quad i \in \mathbf{E}, \quad t = 1, \ldots.
$$

Dan geldt

$$
y_i^t = \sup_{\mathbf{R}} \sum_j \mathsf{P}_i^{\mathbf{R}}(X_t = j) = \sum_j \mathsf{P}_i^{\mathbf{R}_t}(X_t = j),
$$

met $\mathbf{R}^t = (f^{t-1}, \ldots, f^2, f^1, f^1, \ldots)$. Bovendien geldt $y_i^1 \geq \cdots \geq y_i^t \geq y_i^{t+1} \geq \cdots$ en $\lim_{t\to\infty} y^t = 0$.

Bewijs.

De eerste stappen van het bewijs gaan zoals die van lemma 3.2.2 uit [2]. Nu gaan we met inductie bewijzen dat $y^{t-1} \geq y^t$. Voor $t = 1$ geldt: $y^1 \geq y^2$, want

$$
\begin{aligned}
y_i^2 &= \max_a \sum p_{iaj} y_j^1 \\
&= \max_a \sum p_{iaj} \cdot 1 \\
&\leq 1 \\
&= y_i^1.
\end{aligned}
$$

Stel nu $y^1 \geq \cdots \geq y^t$, dan

$$\begin{aligned} y_i^{t+1} &= \sum_j p_{if^t(i)j} y_j^t \\ &\leq \sum_j p_{if^t(i)j} y_j^{t-1} \\ &\leq \max_a \sum_j p_{iaj} y_j^{t-1} \\ &= y_i^t. \end{aligned}$$

We weten nu dat de y^t een monotoon niet stijgende rij vormen met limiet y^*. Stel $y^* \neq 0$. Als f^* een limietpunt is van de rij f^t, zeg $f^{t_k} \to f^*$, dan geldt dat

$$y_i^* = \sum_j p_{if^*(i)j} y_j^*.$$

Kies nu strategie $\mathbf{R}^t = (f^*, \ldots, f^*, f^*, f^*, \ldots)$ en i zodat $y_i^* \neq 0$. Dan geldt vanwege de bovenstaande gelijkheid dat

$$y_i^* = \sum_j \mathsf{P}_i^{\mathbf{R}}(X_t = j) y_j^*.$$

zodat

$$\sum_{t=1}^T \sum_j \mathsf{P}_i^{\mathbf{R}}(X_t = j) y_j^* = T y_i^* \to \infty \quad \text{voor} \quad T \to \infty.$$

Dus

$$\sum_{t=1}^T \sum_j \mathsf{P}_i^{\mathbf{R}}(X_t = j) \geq \sum_{t=1}^T \sum_j \mathsf{P}_i^{\mathbf{R}}(X_t = j) y_j^* \to \infty.$$

En dit laatste is in tegenspraak met de transiëntie-eigenschap van het model.
□

Gevolg 4.3 *Onder de condities van Lemma 4.2 geldt dat* $\sum_t y^t < \infty$.

Bewijs. Omdat $1 \geq y^t \downarrow 0$, $t \to \infty$, en omdat $|\mathbf{E}| < \infty$, geldt dat $a^t = \max_i y_i^t \downarrow 0$, $t \to \infty$. Kies nu T zódat $\max_i y_i^T \leq 1/2$ en laat $\eta = \max_{t \leq T, i \in \mathbf{E}} y_i^t$.

Voor $s \geq 2$

$$\begin{aligned}
y_i^{sT} &= \sum_j \mathsf{P}_i^{\mathbf{R}_{sT}}(X_{sT} = j) \\
&= \sum_k \mathsf{P}_i^{\mathbf{R}_{sT}}(X_{(s-1)T} = k) \cdot \sum_j \mathsf{P}_k^{\mathbf{R}_T}(X_T = j) \\
&\leq \sum_k \mathsf{P}_i^{\mathbf{R}_{sT}}(X_{(s-1)T} = k) y_k^T \\
&\leq \frac{1}{2} \sum_k \mathsf{P}_i^{\mathbf{R}_{sT}}(X_{(s-1)T} = k) \\
&\leq \frac{1}{2} \sum_k \mathsf{P}_i^{\mathbf{R}_{(s-1)T}}(X_{(s-1)T} = k) = \frac{1}{2} y_i^{(s-1)T}.
\end{aligned}$$

Voor $s = 2$ volgt direct dat $\max_i y_i^{sT} \leq (1/2)^2$. Met volledige inductie volgt nu dat $\max_i y_i^{sT} \leq (1/2)^s$.

Voor $t < T$ geldt analoog

$$\begin{aligned}
y_i^{sT+t} &= \sum_k \sum_j \mathsf{P}_i^{\mathbf{R}_{sT+t}}(X_{sT} = k) \cdot \sum_j \mathsf{P}_k^{\mathbf{R}_t}(X_t = j) \\
&\leq \sum_k \sum_j \mathsf{P}_i^{\mathbf{R}_{sT+t}}(X_{sT} = k) \cdot y_k^t \\
&\leq \sum_k \sum_j \mathsf{P}_i^{\mathbf{R}_{sT+t}}(X_{sT} = k) \cdot \eta \\
&\leq \eta \cdot \frac{1}{2^s}.
\end{aligned}$$

Hieruit volgt dat

$$\sum_{t \geq 1} y_i^t \leq \sum_{s \geq 0} \eta \cdot T \cdot 1/2^s = \eta \cdot T$$

$\square$

Stelling 4.4 *Voor een transiënte* MBK *met eindige toestands- en actieruimtes geldt dat de waardevector* $v = \sup_{\mathbf{R}} v(\mathbf{R})$ *eindig is en de unieke oplossing is van het stelsel*

$$x_i = \max\{r_{ia} + \sum_j p_{iaj} x_j\}. \tag{1}$$

Laat f *de beslisregel zijn die de maximaliserende acties kiest in 1, d.w.z* $r_{if(i)} + \sum_j p_{if(i)j} x_j = \max\{r_{ia} + \sum_j p_{iaj} x_j\}$*, dan is* f^∞ *een optimale strategie.*

Bewijs: Laat $r = \max_{a \in \mathbf{A}(i), i \in \mathbf{E}} r_{ia}$. Dan geldt voor een willekeurige strategie $\mathbf{R}' = (\pi^1, \pi^2, \ldots)$ dat

$$\begin{aligned}
v_i(\mathbf{R}') &= \sum_{t \geq 1} \sum_{k,a} \mathsf{P}_i^{\mathbf{R}'}(X_t = k, Y_t = a) r_{ka} \\
&\leq r \sum_{t \geq 1} \sum_{k,a} \mathsf{P}_i^{\mathbf{R}'}(X_t = k) \pi^t(a) \\
&\leq r \sum_{t \geq 1} \sum_{k} \mathsf{P}_i^{\mathbf{R}'}(X_t = k) \\
&\leq r \sum_{t \geq 1} \sup_{\mathbf{R}} \sum_{k} \mathsf{P}_i^{\mathbf{R}}(X_t = k) \\
&\leq r \sum_{t \geq 1} y_i^t.
\end{aligned}$$

Vanwege Gevolg 4.3 geldt nu $v = \sup_{\mathbf{R}} v(\mathbf{R}) \leq r \sum_{t \geq 1} y^t < \infty$.

Laat $\mathbf{R} = (\pi^1, \pi^2, \ldots)$ en $\mathbf{R}^1 = (\pi^2, \pi^3, \ldots)$, dan

$$\begin{aligned}
v_i(\mathbf{R}) &= \sum_a \pi_{ia}^1 (r_{ia} + \sum_j p_{iaj} v_j(\mathbf{R}^1)) \\
&\leq \max_a \{r_{ia} + \sum_j \mathsf{P}_{iaj} v_j(\mathbf{R}^1)\} \\
&\leq \max_a \{r_{ia} + \sum_j \mathsf{P}_{iaj} v_j\}.
\end{aligned}$$

Dus ook

$$v_i = \sup_{\mathbf{R}} v_i(\mathbf{R}) \leq \max_a \{r_{ia} + \sum_j \mathsf{P}_{iaj} v_j\}.$$

Kies nu a_i zodat

$$r_{ia_i} + \sum_j \mathsf{P}_{ia_ij} v_j = \max_a \{r_{ia} + \sum_j \mathsf{P}_{iaj} v_j\}.$$

Kies verder voor elke $j \in \mathbf{E}$ een strategie $\mathbf{R}_j$ willekeurig en laat

$$\mathbf{R} = \{\pi^1, \mathbf{R}_j \,\text{als}\, X_2 = j\}.$$

Dan geldt

$$v_i \geq v_i(\mathbf{R}) = r_{ia_i} + \sum_j \mathsf{P}_{ia_ij} v_j(\mathbf{R}_j).$$

Omdat R_j willekeurig gekozen is en $i \in \mathbf{E}$ geldt

$$\begin{aligned}
v_i &\geq r_{ia_i} + \sum_j \mathsf{P}_{ia_ij} \sup_{\mathbf{R}_j} v_j(\mathbf{R}_j) \\
&= r_{ia_i} + \sum_j \mathsf{P}_{ia_ij} v_j.
\end{aligned}$$

Dan geeft de combinatie van de twee ongelijkheden dat v een oplossing is van de optimaiteitsvergelijking en er volgt ook direct dat $v = v(f^\infty)$.

Nu moet ik nog bewijzen dat de oplossing uniek is. Stel er zijn twee oplossingen; v met bijbehorende strategie f en x met bijbehorende strategie g. Dan

$$\begin{aligned} v_i - x_i &= r_{if(i)} + \sum_j p_{if(i)j} v_j - r_{ig(i)} - \sum_j p_{ig(i)j} x_j \\ &\leq r_{if(i)} + \sum_j p_{if(i)j} v_j - r_{if(i)} - \sum_j p_{if(i)j} x_j \\ &\leq \sum_j p_{if(i)j} [v_j - x_j]. \end{aligned}$$

Dus

$$\begin{aligned} v - x \quad &\leq \quad P(f)[v-x] \\ &\leq \cdots \leq \quad P(f)P(f)[v-x] \\ &= \quad P^2(f)[v-x] \\ &\leq \quad P^t(f)[v-x] \to 0 \quad \text{voor} \quad t \to \infty. \end{aligned}$$

En daaruit volgt dan

$$v - x \leq 0 \Longrightarrow v \leq x.$$

Op dezelfde wijze volgt

$$v - x \geq P(g)[v-x] \geq P^t(g)[v-x] \geq 0.$$

Dus is $v \geq x$ en derhalve moeten v en x gelijk zijn en is de oplossing van (1) uniek. En dat vervolledigt het bewijs. $\square$

5 Successieve approximatie

Een methode die geschikt is om het stelsel uit stelling 4.2 op te lossen is successieve approximatie. Bij deze methode ga je uitgaande van een willekeurige beginvector v^0, de waardevector v en een optimale stationaire strategie f^∞ benaderen.

Stelling 5.1 *Laat de rij $\{v^n\}_{n=1}^{\infty}$ gedefinieerd door:*

$$\begin{cases} v^0 \in R^N, & \textit{willekeurig} \\ v_i^{n+1} = \max_a\{r_{ia} + \sum p_{iaj} v_j^n\}, & n\textit{=1,2,....} \end{cases}$$

Dan convergeert v^n naar v en f^n naar f^∞, waarbij f^n de beslisregel is die de maximaliserende acties kiest in de n-de iteratiestap.

Bewijs. Neem eerst aan dat $v^0 = 0$. Eerst bewijzen we met volledige inductie dat $v^0 \leq .. \leq v^n \leq v^{n+1} \leq .. \leq v$. Ten eerste geldt dat $v^1 \geq v^0$, omdat $v_i^1 = \max_a r_{ia} \geq 0 = v_i^0$. We nemen nu aan dat $v^0 \leq .. \leq v^n$ voor $n \leq N$ en we zullen bewijzen dat $v^N \leq v^{N+1}$. Bekijk het verschil

$$\begin{aligned} v_i^{N+1} - v_i^N &\geq r_i f_i^N + \sum \mathsf{P}_{if_i^N j} v_j^N - (r_i f_i^N + \sum \mathsf{P}_{if_i^N j} v_j^{N-1}) \\ &= \sum \mathsf{P}_{if_i^N j}(v_j^N - v_j^{N-1}) \\ &\geq 0. \end{aligned}$$

Deze laatste ongelijkheid geldt vanwege de inductieveronderstelling dat $v^N \geq v^{N-1}$. Als we dan nu strategie $\mathbf{R}$ spelen, die er als volgt uitziet: $\mathbf{R} = \{f^N, f^{N-1}, .., f^1, f^1, f^1, ...\}$, dan volgt dat $v^N \leq v(\mathbf{R}) \leq v$, omdat we niet-negatieve opbrengsten veronderstellen. We weten nu dus dat de v^N een monotoon stijgende rij vormen, begrensd door v. Deze rij heeft dus een limiet, zeg v', met $v' \leq v$.

Laat nu f' een limietpunt zijn van de rij f^n, dan is deze de limiet van een deelrij f^{n_k}, dus $f' = \lim_{k\to\infty} f^{n_k}$. Vanwege de keuze van f^{n_k} geldt

$$v_i^{n_k} = r_{if_i^{n_k}} + \sum_j \mathsf{P}_{if_i^{n_k} j} v_j^{n_{k-1}} \geq r_{ia} + \sum \mathsf{P}_{iaj} v_j^{n_{k-1}}, \quad \forall a \in \mathbf{A}(i).$$

Als we dan nu de limiet $k \to \infty$ nemen, wat mag omdat de sommatie over j eindig is, dus mogen we limiet en som verwisselen, dan volgt:

$$v_i' = r_{if_i'} + \sum_j \mathsf{P}_{if_i' j} v_j' \geq r_{ia} + \sum_j \mathsf{P}_{iaj} v_j' \quad \forall a \in \mathbf{A}(i).$$

En daaruit volgt dus dat:

$$v_i' = \max_a\{r_{ia} + \sum_j \mathsf{P}_{iaj} v_j'\}.$$

En omdat v de unieke oplossing is, volgt nu dat $v' = v$ en ${f'}^\infty$ is een optimale strategie voor elke limietpunt f' van de rij $\{f^n\}$. Voor $v^0 = 0$ completeert dit het bewijs.

Nu het geval $v^0 \geq 0$. Laat $\{v^n\}_{n=1}^\infty$ een rij uitkomsten van het algoritme met bijbehorende beslisregels f^n voor $v^0 \geq 0$. Laat evenzo $\{(v')^n\}_{n=1}^\infty$ een rij uitkomsten van het algoritme met bijbehorende beslisregels ${f'}^n$ voor $v^0 = 0$. We hadden al dat ${v'}_i^1 = \max_a r_{ia} = r_{(f')^1}$. Om te bewijzen dat de stelling ook geldt voor willekeurige v^0 is het voldoende te bewijzen dat:

$$ {v'}^n \leq v^n \leq {v'}^n + P(f^n) \cdots P(f^1) v^0. $$

Vanwege de transiënte eigenschap convergeert $\sup_{f^n,\ldots,f^1} P(f^n) \cdots P(f^1) v^0$ naar 0 voor $n \to \infty$.

Met volledige inductie zullen we bewijzen dat

$$ {v'}^n + P(f^n) \cdots P(f^1) v^0 \geq v^n. $$

Voor $n = 1$ geldt:

$$ {v'}^1 + P(f^1) v^0 \geq r_{(f')^1} + P(f^1) v^0 \geq v^1. $$

Nu de inductiestap:

$$ \begin{aligned} {v'}^n + P(f^n) P(f^{n-1}) \cdots P(f^1) v^0 &\geq r_{f^n} + P(f^n) {v'}^{n-1} + P(f^n) \cdots P(f^1) v^0 \\ &= r_{f^n} + P(f^n) \big[{v'}^{n-1} \\ &\qquad + P(f^{n-1}) \cdots P(f^1) v^0 \big] \\ &\geq r_{f^n} + P(f^n) v^{n-1} = v^n. \end{aligned} $$

Voor de ondergrens bekijken we het verschil $v^n - {v'}^n$. Voor $n = 1$ geldt:

$$ v^1 - {v'}^1 \geq r_{(f')^1} + P((f')^1) v^0 - r_{(f')^1} = P((f')^1) v^0 \geq 0. $$

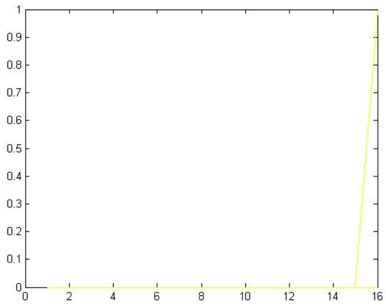
Dus $v^1 \geq {v'}^1$, Nu de inductiestap:

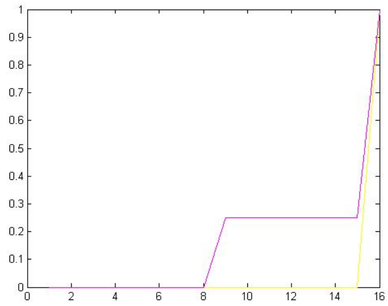
$$ \begin{aligned} v^n - {v'}^n &\geq r_{(f')^{n-1}} + P((f')^{n-1}) v^{n-1} - r_{(f')^{n-1}} - P((f')^{n-1}) (v')^{n-1} \\ &= P((f')^{n-1}) (v^{n-1} - (v')^{n-1}) \\ &\geq 0. \end{aligned} $$

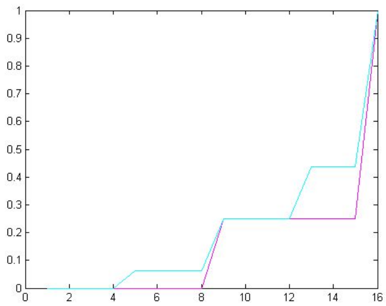
Dus $v^n \geq {v'}^n$. Voor algemene v^0 gaat het analoog. □

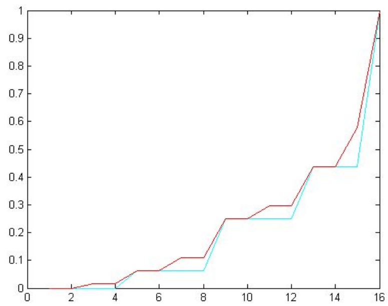
Om een beter beeld te krijgen van het bovenstaande algoritme is hier een voorbeeld van hoe de benaderingen van de waardevector er tijdens het algoritme in de achtereenvolgende iteratiestappen uitziet. Elk van de plaatjes representeert twee opeenvolgende iteratiestappen van het algoritme om te laten zien hoe het algoritme naar de oplossing convergeert. De plaatjes zijn gemaakt voor het geval dat $N = 16$, p=0.25 en beginvector $v^0 = 0$.

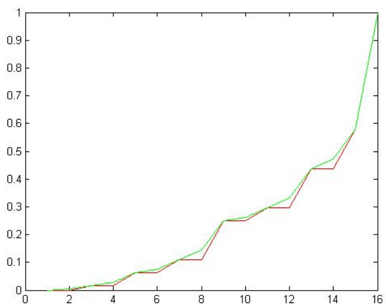
In het onderstaande tabelletje staat voor elke iteratie de gebruikte kleur. Deze kleuren gebruik ik ook in alle volgende plaatjes met waardevectoren.
1e iteratie geel, 2e iteratie rose
3e iteratie cyaan, 4e iteratie rood
5e iteratie groen, 6e iteratie blauw
7e iteratie zwart, 8e iteratie geel
9e iteratie rose, 10e iteratie cyaan

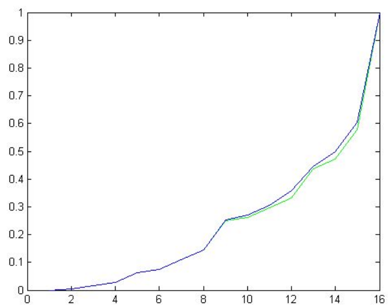


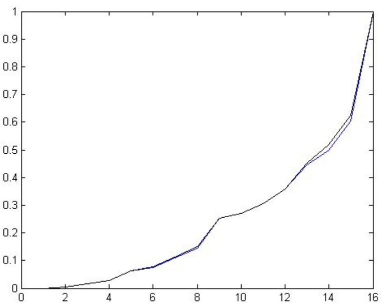


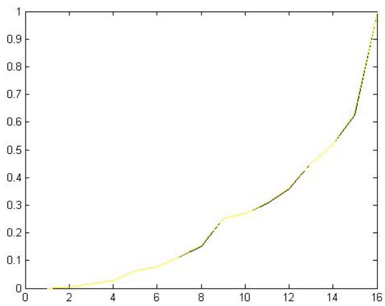


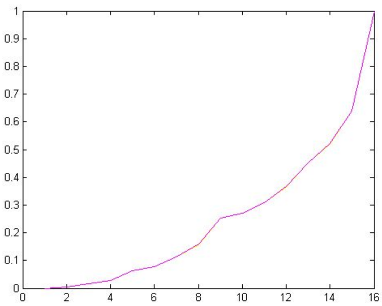







1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0
2
4
6
8
10
12
14
16


1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0
2
4
6
8
10
12
14
16


1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0
2
4
6
8
10
12
14
16

6 Strategieën

Nu we dus een methode hebben om een waardevector uit te rekenen en optimale strategieën te vinden, kunnen we voor het Red en Black model de optimale strategie en bijbehorende waardevector berekenen. Bij de overgangskansen moet ik echter onderscheid maken tussen 3 gevallen, namelijk $p \leq 1/2$, $p = 1/2$ en $p \geq 1/2$.

Stelling 6.1 *Voor $p = 1/2$ is elke stationaire strategie f^∞ optimaal.*

Bewijs: Neem een f^∞ willekeurig, dan is $v(f^\infty)$ de unieke oplossing van:

$$\begin{cases} x_0 &= 0 \\ x_i &= 1/2\{x_{i+a} + x_{i-a}\} \\ x_N &= 1. \end{cases}$$

Het is duidelijk te zien dat de oplossing $x_i = i/N$ voldoet en daaruit volgt dat de waardevector $v_i(f^\infty) = i/N$, onafhankelijk van strategie f^∞. Dus $v_i = i/N$, $0 \leq i \leq N$ is de waardevector en f^∞ is een optimale strategie. $\square$

Stelling 6.2 *Voor $p > 1/2$ blijkt dat timide spel, d.w.z. $f_t(i) = 1$, $\forall i$, optimaal is.*

Bewijs: De waardevector $v(f_t)$ van timide spel is de oplossing van het stelsel:

$$\begin{cases} x_0 &= 0 \\ x_i &= px_{i+1} + qx_{i-1} \, voor \, 1 \leq i \leq N-1 \\ x_N &= 1. \end{cases}$$

Omdat $x_1 = qx_0 + px_2 = px_2$ is dit stelsel door eenvoudige substituties op te lossen. De oplossing is zelfs in deze vorm te schrijven: $x_i = \frac{1-(q/p)^i}{1-(q/p)^N}$. Voor $1 \leq i \leq N-1$ hebben we:

$$\begin{aligned} px_{i+1} + qx_{i-1} &= p \cdot \frac{1-(q/p)^{i+1}}{1-(q/p)^N} + q \cdot \frac{1-(q/p)^{i-1}}{1-(q/p)^N} \\ &= \frac{1-(q/p)^{i-1}(\frac{q^2}{p}+q)}{(1-(q/p)^N)} = x_i, \end{aligned}$$

immers $\frac{q^2}{p} + q = \frac{q}{p}(q+p) = q/p$.

Om nu te bewijzen dat f_t een optimale strategie is, is het voldoende aan te tonen dat $v_i(f_t) \geq px_{i+a} + qx_{i-a}$, $\forall i \in \mathbf{E}, a \in A(i)$. Het is dus voldoende aan te tonen dat

$$(q/p)^i \leq p(\frac{q}{p})^{i+a} + q(\frac{q}{p})^{i-a} \quad \text{oftewel dat} \quad 1 \leq p(\frac{q}{p})^a + q(\frac{q}{p})^{-a}.$$

Laat

$$F(a) = p(\frac{q}{p})^a + q(\frac{q}{p})^{-a},$$

dan geldt:

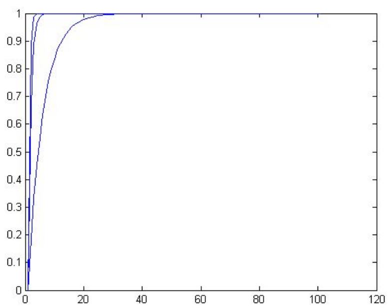
$$\begin{aligned} F(a+1) \geq F(a) &\iff p(\frac{q}{p})^{a+1} + q(\frac{q}{p})^{-a-1} \geq p(\frac{q}{p})^{a} + q(\frac{q}{p})^{-a} \\ &\iff pq^{2a+2} + qp^{2a+2} \geq p^2 q^{2a+1} + q^2 p^{2a+1} \\ &\iff q^{2a+1} + p^{2a+1} \geq pq^{2a} + qp^{2a} \\ &\iff p^{2a}(p-q) \geq q^{2a}(p-q). \end{aligned}$$

In de eennalaatste deel ik door pq. En deze laatste ongelijkheid geldt omdat $p > q$. $\square$

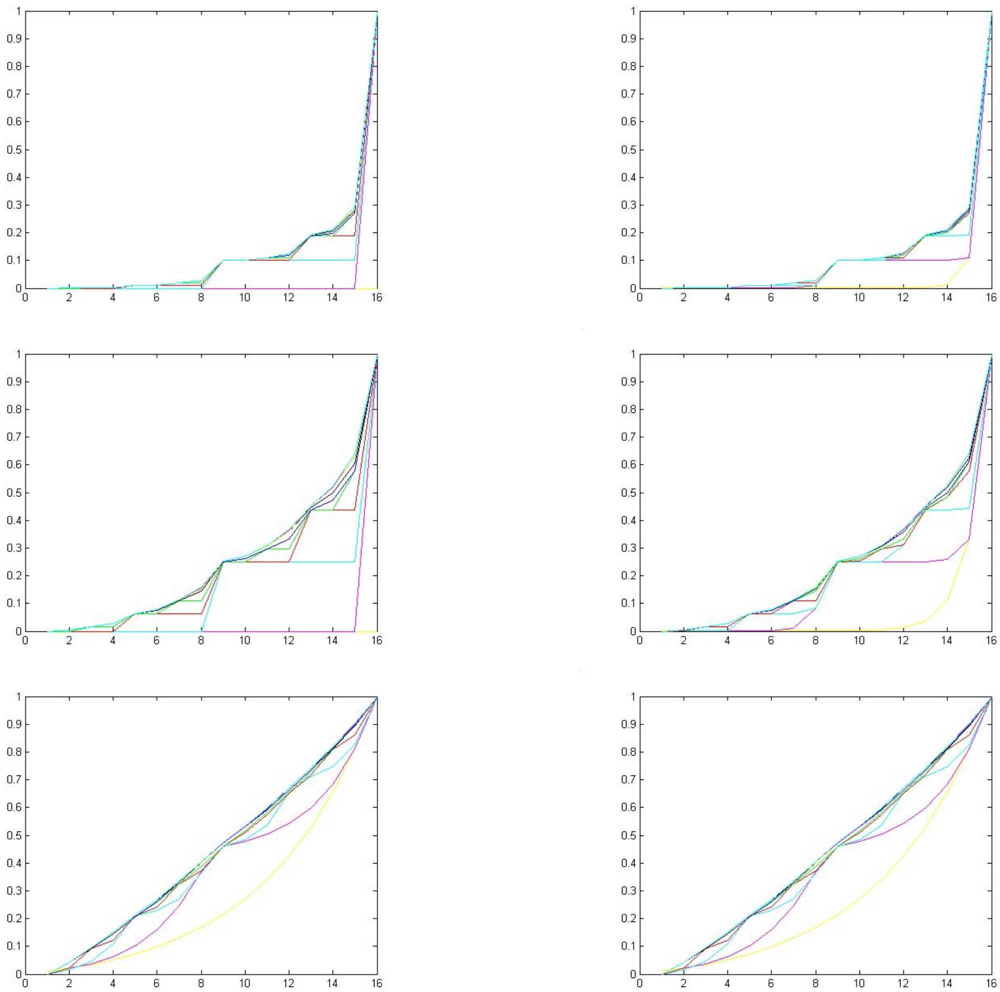
Het geval $p < 1/2$ behandel ik hier niet, omdat er geen analytische uitdrukking voor de waardevector bekend zijn. In [4] staat wel een bewijs dat stout spel, dat wil zeggen dat je steeds zo hoog mogelijk inzet, maar ook niet meer dan precies genoeg om het streefbedrag te halen, de beste strategie is. Dit bewijs is echter vrij gecompliceerd en geeft geen exacte oplossing voor de waardevector. Ik heb wel geprobeerd om daar zelf wat over te bewijzen, maar dat is niet gelukt.

7 Toepassingen

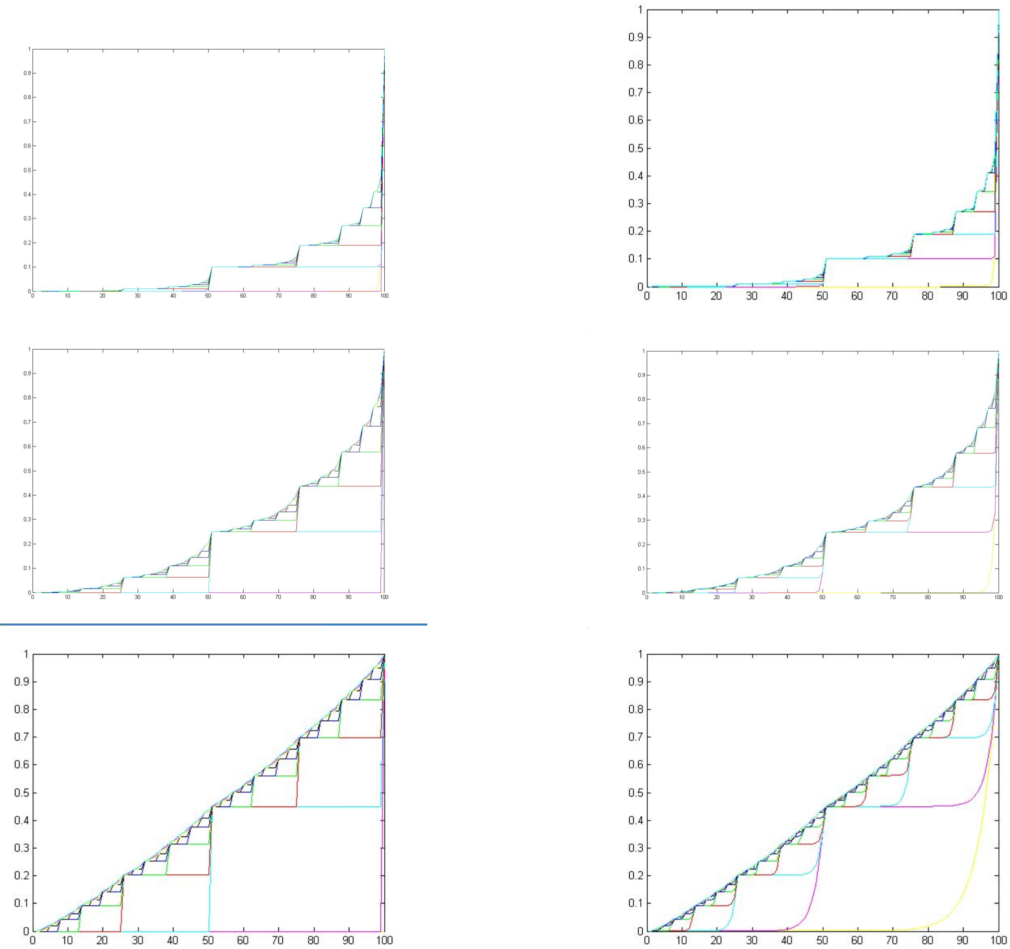
In dit hoofdstuk wil ik gebruik makend van de voorgaande theorie een aantal voorbeelden behandelen om de theorie inzichtelijker te maken. Als toestandsruimten wil ik onderscheid maken tussen twee gevallen, namelijk $\mathbf{E} = \{0, 1, ..., 99\}$ en $\mathbf{E} = \{0, 1, ..., 15\}$. Voor deze twee gevallen wil voor verschillende winstkansen bekijken hoe de waardevector eruit ziet. Voor $p \geq 1/2$ kan dit analytisch en voor $p \leq 1/2$ doe ik dit met de methode van successieve appoximatie. Verder is het bij het gebruik van de methode van successieve appoximatie interessant om de startvector v^0 te variëren, aangezien dit snellere convergentie op zou kunnen leveren en misschien ook een hint om een eenvoudiger bewijs te formuleren dat stout spel optimaal is voor $p \leq 1/2$. Voor $n = 16$ en $n = 100$ heb ik de waardevectoren voor p=0.9, p=0.75 en p=0.55 berekend en de resultaten samengevoegd in matlab in de volgende figuren. Van boven naar beneden zijn het de grafieken van respectievelijk p=0.9, p=0.75 en p=0.55.



Interessanter zijn echter de gevallen waar $p \leq 1/2$ aangezien dit de gevallen zijn die je in een casino doorgaans aan zult treffen. Ik heb voor beide toestandsruimten steeds voor drie kansen $p \leq 1/2$, namelijk $p = 0.45$, $p = 0.25$ en $p = 0.1$, een aantal iteraties van het successieve approximatie-algoritme uitgevoerd voor twee verschillende beginvectoren v^0 en de tussenresultaten per iteratie opgeslagen in grafiekjes. In de linkerkolom staan steeds de resultaten met beginvector $v^0 = 0$ en in de rechterkolom de resultaten met als beginvector v^0 de oplossing van timide spel.



De resultaten bij lage winstkans zien er nog niet erg mooi uit, maar naarmate de winstkans groter wordt lijkt er meer structuur in te komen. Opvallend is ook, dat de rechterplaatjes, waarin ik in plaats van een beginvector $v^0 = 0$ als beginvector de oplossing van het timide spel gebruik (Stelling 6.2), sneller convergeert naar de optimale oplossing. Het leek me echter mogelijk dat een 2-macht misschien mooier gestructureerd zou zijn dan een toestandsruimte van andere grootte. Dit mooie fenomeen bleek echter niet alleen van toepassing op toestandsruimten met N een macht van 2, zoals wel te zien is uit de volgende serie plaatjes en vooral het geval $p = 45/100$.



Maar ondanks het feit dat er meer structuur in lijkt te komen voor een andere beginvector in het algoritme, blijft het verloop van de waardevector vrij grillig, wat de mogelijkheden om een mooie exacte formule voor de oplossing te vinden ernstig beperkt. Voor $p < 1/2$ zullen we ons tevreden moeten stellen met een successieve approximatie bewijs om de waardevector en optimale strategie te bepalen.

8 Conclusie

In eerste instantie was het doel van het onderzoek om voor het gokmodel de optimale strategieën voor de verschillende winstkansen te berekenen. Nu ga ik aan de hand van de in de inleiding gestelde doelen bespreken wat wel en niet gelukt is. Het formeel formuleren van het gokmodel is gelukt en dit is in de sectie 'Model en Notatie' verwerkt. Vervolgens is het ook gelukt het gokmodel als totale opbrengstenmodel te schrijven, maar omdat er nog wat nadelen zijn wat betreft de eindigheid heb ik dit nog omgeschreven naar een transiënt model. Daarvoor heb ik de bestaande theorie die via een verdisconteerd model ging, omgeschreven naar die voor een transiënt model. Ditzelfde is ook gelukt voor de theorie van de successieve approximatie. Vervolgens heb ik uit [4] de bewijzen bestudeerd over de optimale strategieën van het gokmodel. De bewijzen voor $p \geq 1/2$ heb ik opgenomen in de sectie 'Strategieën', maar het bewijs voor $p < 1/2$ heb ik hier niet in opgenomen. Dit bewijs is namelijk erg gecompliceerd en geeft ook geen exacte uitdrukking voor de waardevector. Ik heb zelf geprobeerd om een eenvoudiger bewijs te vinden, maar dit is niet gelukt. Het is echter wel duidelijk wat de optimale strategie is. De optimale strategieën voor de verschillende winstkansen zijn dan:

$$A(i) = \begin{cases} 1, & \text{als } p \geq 1/2 \\ \min\{i, N-i\}, & \text{als } p \leq 1/2 \\ \text{willekeurig}, & \text{als } p = 1/2 \end{cases}$$

Om toch voor $p < 1/2$ nog onderzoek te kunnen doen heb ik de in de appendix opgenomen matlabcode gemaakt om de waardevector via de methode van successieve approximatie te benaderen. Door deze plaatjes werd het duidelijk waarom het niet eenvoudig zal zijn een analytische uitdrukking voor de waardevector te vinden, aangezien het verloop van de waardevector erg grillig is. Al met al heb ik eigenlijk alle gestelde doelen gehaald dus mag ik stellen dat het een geslaagd onderzoek is geweest.

Referenties

[1] E.L. Dubins and L.J. Savage (1965), *How to gamble if you must : Inequalities for Stochastic Processes.* McGraw-Hill, New York.

[2] L.C.M. Kallenberg (1980) *Linear programming and finite Markovian control problems.* Mathematical Centre Tracts **148**, CWI, Amsterdam.

[3] L.C.M. Kallenberg (2004), *Dictaat bij college Besliskunde 2.*

[4] L.C.M. Kallenberg (1994), *Markov Decision Theory, Lecture notes.*

9 Appendix

Voor geïnteresseerden heb ik de matlabcode bijgevoegd die ik gebruikt heb om de grafiekjes te maken.

gamble.m

```
function [v1,v2,u,v]=gamble(n,p,x)
% berekent optimale strategie en waarde functie in
% een successieve approximatie stap
v1=[]; v2=[]; v3=[]; v1=[v1 0]; v2=[v2 0];
 for k=2:n
   for a=1:min(k-1,n+1-k)
     v3=[v3 ((1-p)*x(k-a)+p*x(k+a))];
   end
  v1=[v1 max(v3)];
  hits=find(v3>=max(v3));
  v2=[v2 max(hits)];
  v3=[];
 end
v2=[v2 0]; v1=[v1 1]; u=zeros(1,n); for i=1:(n)
    u(i)=v1(i+1)-v1(i);
end v=zeros(1,floor(n/2)); for i=1:floor(n/2)
    v(i)=x(i)+x(n+2-i);
end
}
```

Gebruik: `x` moet aangemaakt worden, bijv. `x=zeros(16);`. Dan de aanroep: `function [v1,v2,u,v]=gamble(n,p,x)` met voor `n` en `p` waarden ingevuld.

itereren.m

```
function [k1,k2,k3,k4,k5,k6,k7,k8,k9,k0]=itereren(n,p,x)
k1=zeros(n); k2=zeros(n); k3=zeros(n); k4=zeros(n); k5=zeros(n);
k6=zeros(n); k7=zeros(n); k8=zeros(n); k9=zeros(n); k0=zeros(n);

k1=x; [k2,y,u,v]=gamble(n,p,k1);

[k3,y,u,v]=gamble(n,p,k2);

[k4,y,u,v]=gamble(n,p,k3);

[k5,y,u,v]=gamble(n,p,k4); [k6,y,u,v]=gamble(n,p,k5);
[k7,y,u,v]=gamble(n,p,k6); [k8,y,u,v]=gamble(n,p,k7);
[k9,y,u,v]=gamble(n,p,k8); [k0,y,u,v]=gamble(n,p,k9);
```

```
plot(k1,'y'); hold on plot(k2,'m'); hold on plot(k3,'c'); hold on
plot(k4,'r'); hold on plot(k5,'g'); hold on plot(k6,'b'); hold on
plot(k7,'k'); hold on plot(k8,'y'); hold on plot(k9,'m'); hold on
plot(k0,'c');
```

Gebruik: `x` moet aangemaakt worden, bijv. `x=zeros(16);`. Dan de aanroep `function [k1,k2,k3,k4,k5,k6,k7,k8,k9,k0]=itereren(n,p,x)` met voor `n` en `p` waarden ingevuld.