februar 1968

DYNAMISK STYRING

av Tore Schweder

I Dynamisk styring av markovkjede
med diskret tid ..... 1.

Trinnverdi iterasjon 8

Verdidifferens-iterasjon 16

Eksempler 23

II Dynamisk styring av markovprosess
med kontinuerlig tid 31

Verdidifferens-iterasjon 37

Denne stensilen er laget for Erling Sverdrup og under hans oppsyn.

Både teorien og eksemplene bygger på Ronald A. Howard: DYNAMIC PROGRAMMING AND MARKOV PROCESSES. Noen av eksemplene er tatt direkte fra denne boka.

Som viktig verktøy i utviklingen er genererende funksjoner og Laplace-transformasjon brukt. Disse metodene er utviklet i stensilen OPERATORMETODER I SANNSYNLIGHETS-REGNING av Erling Sverdrup.

Generelt om dynamisk styring.

Mange situasjoner i dagliglivet kan beskrives som markovprosesser. Ofte kan vi velge hvilken situasjon vi vil være i, vi velger hvorledes markovprosessen skal være, vi styrer markovprosessen. Det valget vi gjør er basert på en vurdering av hvilken situasjon som er nyttigst eller har størst verdi. Dynamisk styring eller dynamisk programmering er betegnelsen på de metodene en kan bruke til å finne det mest verdifulle i en mengde operasjonelle systemer.

Vi skal utelukkende beskjeftige oss med dynamisk styring av markovprosesser, det vil si metoder til å finne den mest verdifulle i en mengde markovprosesser.

I DISKRET TID

Gitt en markovprosess $\langle X_n \rangle_{n=0}^{\infty}$ med endelig tilstandsrom $\{1,2,\ldots,N\}$ og med én rekurent klasse. Vi skal si at $\langle X_n \rangle$ er en prosess med fortjeneste hvis det er tilknyttet en fortjenestematrise R og prosessen tjener r_{ij} når den gjør en overgang fra tilstand i til tilstand j.

Hvis $X_o = i$ er $v_i(n)$ den forventede gevinsten etter n overganger (på tidspunkt n). Vi skal kalle $v_i(n)$ n-<u>trinnverdien</u>. Når P er overgangsmatrisen har vi

$$v_i(n) = \sum_{j=1}^{N} Pr(X_1 = j | X_o = i)(r_{ij}+v_j(n-1))$$

som gir

$$(1) \qquad v_i(n) = \sum_{j=1}^{N} P_{ij}r_{ij} + \sum_{j=1}^{N} P_{ij}v_j(n-1); \quad i = 1,\ldots,N$$

La $v(n)$, q_i og q være definert slik

$$v(n) = \begin{bmatrix} v_1(n) \\ \vdots \\ v_N(n) \end{bmatrix}, \quad q_i = \sum_{j=1}^{N} P_{ij}r_{ij} \quad \text{og} \quad q = \begin{bmatrix} q_1 \\ \vdots \\ q_N \end{bmatrix}$$

Ligning (1) blir på vektorform

$$(2) \qquad v(n) = q+Pv(n-1)$$

Ved å la $\varphi_i(s)$ være den genererende funksjon til $v_i(n)$, $\varphi_i(s) = \sum_{n=0}^{\infty} v_i(n)s^n$, og $\underline{\Phi}$ den genererende funksjon til v, får vi av (2)

$$\frac{1}{s}(\underline{\Phi}(s)-\underline{\Phi}(0)) = \frac{1}{1-s}q+P\underline{\Phi}(s)$$

$$(I-sP)\underline{\Phi}(s) = \frac{s}{1-s}q+v(0)$$

$$\underline{\Phi}(s) = (I-sP)^{-1}(\frac{s}{1-s}q+v(0)$$

Vi ser at for å finne $\underline{\Phi}(s)$, må vi bestemme $(I-sP)^{-1}$. Hovedoppgaven blir dermed å bestemme $\underline{\Pi}(s)$ som er den genererende funksjonen for $p(n) = p(0)\cdot P^n$ der $p_i(n) = Pr(X_n = i)$.

I "Operatormetoder i sannsynlighetsregning" ble $(I-sP)^{-1}$ drøftet, og en fant

$$(I-sP)^{-1} = \frac{1}{1-s}S+\underline{\tau}(s)$$

der S er matrisen bestående bare av stasjonærfordelingen $\underset{\sim}{p}$, $S = \begin{bmatrix} \underset{\sim}{p} \\ \vdots \\ \underset{\sim}{p} \end{bmatrix}$. $\underline{\tau}(s)$ er genererende funksjon for en sekvens $\langle t_n \rangle$ som går geometrisk mot null. Følgelig får vi

$$\underline{\Phi}(s) = \frac{s}{(1-s)^2}Sq + \frac{1}{1-s}Sv(0) + \frac{s}{1-s}\underline{\tau}(s)q + \underline{\tau}(s)v(0)$$

Her har vi

$$\frac{s}{(1-s)^2}Sq = \sum_{n=0}^{\infty} ns^n Sq$$

og $$\frac{s}{1-s}\bar{\tau}(s) = \frac{1}{1-s}\bar{\tau}(s) - \bar{\tau}(s) = \sum_{n=0}^{\infty} s^n \sum_{m=0}^{n} t_m - \bar{\tau}(s)$$

$$= \sum_{n=0}^{\infty} s^n \sum_{m=0}^{n-1} t_m$$

Altså $v(n) = nSq + q\sum_{m=0}^{n-1} t_m + Sv(0)$

Nå er $\sum_{m=0}^{n-1} t_m = \bar{\tau}(1) + \varepsilon_n$ hvor ε_n går geometrisk mot null. Dermed har vi

$$v(n) = nSq + \bar{\tau}(1)q + Sv(0) + \varepsilon_n$$

$v(n)$ er dèt markovkjeden har tjent etter n trinn. Det er markovkjedens verdi etter n trinn. Når n vokser, vokser også $v(n)$ over alle grenser (hvis $Sq \neq 0$). Jo fortere $v(n)$ vokser, dess mer verdifull er markovkjeden. Hvis vi er interessert i en markovkjede som går svært lenge, er det naturlig å måle dens verdi ved Sq.

Vi skal definere <u>markovkjedens gevinst</u> g ved

$$\begin{bmatrix} g \\ \vdots \\ g \end{bmatrix} = Sq = \lim_{n\to\infty} \frac{v(n)}{n} .$$

Noe vi kan merke oss er at fortjeneste-matrisen R ikke har noen betydning uten gjennom q.

Eksempel

En drosjesjåfør opererer i et distrikt med tre byer A, B og C. Han får turer innen og mellom byene. Hvis han på tidspunkt n er i A, er $X_n=1$, i B $X_n=2$ og i C $X_n=3$. Drosjen har fortjeneste-matrise .

$$R = \begin{bmatrix} 10 & 4 & 8 \\ 14 & 0 & 18 \\ 10 & 2 & 8 \end{bmatrix} \quad \text{og overgangsmatrise} \quad P = \begin{bmatrix} \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{2} \end{bmatrix}$$

Dette gir $q = \begin{bmatrix} 8 \\ 16 \\ 7 \end{bmatrix}$

Vi har før funnet

$$(I-sP)^{-1} = \frac{1}{1-s}\begin{bmatrix} \frac{2}{5} & \frac{1}{5} & \frac{2}{5} \\ \frac{2}{5} & \frac{1}{5} & \frac{2}{5} \\ \frac{2}{5} & \frac{1}{5} & \frac{2}{5} \end{bmatrix} + \frac{1}{1-\frac{s}{4}}\begin{bmatrix} \frac{1}{2} & 0 & -\frac{1}{2} \\ 0 & 0 & 0 \\ -\frac{1}{2} & 0 & \frac{1}{2} \end{bmatrix}$$

$$+ \frac{1}{1-(\frac{-s}{4})}\begin{bmatrix} \frac{1}{10} & -\frac{1}{5} & \frac{1}{10} \\ -\frac{2}{5} & \frac{4}{5} & -\frac{2}{5} \\ \frac{1}{10} & -\frac{1}{5} & \frac{1}{10} \end{bmatrix}$$

Dermed

$$Sq = \begin{bmatrix} 9.2 \\ 9.2 \\ 9.2 \end{bmatrix} \quad \text{og} \quad \overline{C}(1)\cdot q = \frac{1}{75}\begin{bmatrix} 56 & -12 & -44 \\ -24 & 48 & -24 \\ -44 & -12 & 56 \end{bmatrix} \cdot \begin{bmatrix} 8 \\ 16 \\ 7 \end{bmatrix} \approx \begin{bmatrix} -0.7 \\ 5.3 \\ -2 \end{bmatrix}$$

Når drosjesjåføren starter med tom pung på tidspunkt 0, blir $v(0) = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$.

Følgelig blir drosjens forventede fortjeneste i løpet av n reiser hvis han starter i i-te by

$$v_i(n) = n \cdot 9.2 + \begin{cases} -0.7; & i = 1 \\ +5.3; & i = 2 \\ -2 \ ; & i = 3 \end{cases}$$

Drosjens gevinst blir i dette tilfelle $g = 9.2$.

La $p_i^1, p_i^2, \ldots, p_i^{n_i}$ være n_i forskjellige sannsynlighetsfunksjoner for tilstandene $1,2,\ldots,N$. Hvis $X_n = i$, kan vi velge etter hvilken p_i^k neste overgang skal foregå. (p_i^k er linjevektorer : $p_i^k = [p_{i1}^k, \ldots, p_{iN}^k]$). Hvis vi på et tidspunkt n har bestemt oss for å velge $p_i^{j_i}$ som sannsynlighetsfunksjon for neste overgang for $i = 1,2,\ldots,N$, har vi fastlagt en desisjonsfunksjon, eller kort desisjon, d_n som er slik at $d_n(i) = j_i$. La D være mengden av slike desisjonsfunksjoner,

$$D = \{d \mid d(i) \in \{1,2,\ldots,n_i\}, \ i = 1,\ldots,N\}.$$

På tidspunkt n hadde vi altså bestemt oss for å benytte $d_n \in D$ for å bestemme etter hvilken sannsynlighetsfordeling $(n+1)$-te overgang skal skje. Denne overgangen blir da behersket av overgangsmatrisen

$$P_{n+1} = \begin{bmatrix} p_1^{d_n(1)} \\ p_2^{d_n(2)} \\ \vdots \\ p_N^{d_n(N)} \end{bmatrix}$$

Hvis vi har bestemt $d_n \in D$ for $n = 1,2,\ldots,$ da har vi bestemt en strategi $\delta = \langle d_1, d_2, \ldots \rangle$ for markovkjeden. Vi får da bestemt en overgangsmatrise P_n for alle n, og hele sannsynlighetsstrukturen for markovkjeden blir fastlagt.

En strategi er altså en fullstendig beskrivelse av hvilken sannsynlighetsfordeling X_{n+1} skal ha når vi kjenner X_n, $n = 1,2,\ldots$. Mengden av alle δ skal vi kalle Δ .

En strategi δ der $d_n = d$ for $n = 1,2,\ldots$ skal vi kalle en stasjonær strategi. I dette tilfellet har vi

$$P_n = \begin{bmatrix} p_1^d(1) \\ p_2^d(2) \\ \vdots \\ p_N^d(N) \end{bmatrix} = P \quad \text{for alle} \quad n,$$

og markovkjeden får dermed konstant overgangsmatrise.

Til hver alternativ p_i^k hører en fortjenestevektor r_i^k. Hvis $X_n = i$ og $d_n(i) = k$, så tjener prosessen

r_{ij}^k hvis $X_{n+1} = j$. q_i^k er definert som $\sum_{j=1}^{N} r_{ij}^k \cdot p_{ij}^k = q_i^k$.

I vår behandling av dynamisk styring skal vi forutsette at hver desisjon i D gir opphav til en markovkjede med én rekurent klasse.

I boka til Howard er også det tilfellet behandlet at det kan være flere rekurente klasser. Teknikken blir da mer komplisert, men prinsipielt er det samme fremgangsmåte som i 1-kjede tilfellet.

TRINNVERDI ITERASJON

Problemet i dynamisk styring er å finne den beste strategien. Men hva skal menes med den beste strategien? Hvis vår markovkjede bare skal gjøre et lite antall overganger, er vi mer interessert i at disse overgangene er nyttige, enn at man skal tjene mye på lang sikt. Vi skal demonstrere trinnverdi iterasjonsmetoden, som plukker ut den strategien som maximerer trinnverdien $v_i(n)$ for fast n og alle i.

La oss definere $v_i^*(n)$ slik:

$$v_i^*(n) = \max_{\delta \in \Delta} v_i^{\delta}(n)$$

Der $v_i^{\delta}(n)$ er den forventede n-trinnverdien når strategien δ blir fulgt.

Trinnverdi interasjonen skal plukke ut begynnelsen $d_1^*,\ldots,d_n^*$ av en strategi δ^* som gir $v_i^*(n) = v_i^{\delta^*}(n)$; $i = 1,\ldots,N$.

Nå har vi

$$v_i^*(n+1) = \max_{\delta \in \Delta} v_i^{\delta}(n+1) = \max_{\Delta} \left[q_i^{d_1(i)} + \sum_{j=1}^{N} p_{ij}^{d_1(i)} v_j^{\delta}(n) \right]$$

Altså

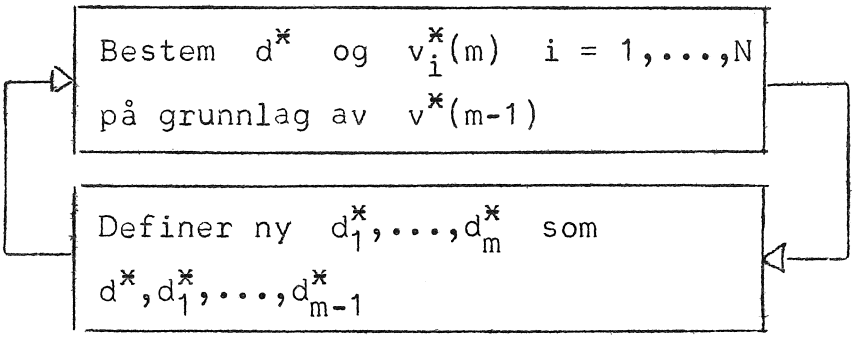
$$v_i^*(n+1) = \max_{k \in \{1,\ldots,n_i\}} \left[q_i^k + \sum_{j=1}^{N} p_{ij}^k v_j^*(n) \right]$$

Denne relasjonen skal vi nyttiggjøre oss. Når $v_j^*(n)$ er kjent for $j = 1,\ldots,N$, skal vi definere d^* ved

$$q_i^{d^*(i)} + \sum_{j=1}^{N} p_{ij}^{d^*(i)} v_j^*(n) = \max_{k \in \{1,\ldots,n_i\}} \left[q_i^k + \sum_{j=1}^{N} p_{ij}^k v_j^*(n) \right]$$

Hvis $d_1^*,\ldots,d_n^*$ gir $v_i^*(n)$ for $i = 1,\ldots,N$, da vil $d^*,d_1^*,\ldots,d_n^*$ gi trinnverdien $v_i^*(n+1)$ for $i = 1,\ldots,N$.

Når man kjenner startbetingelsene $v_j(0)$; $j = 1,\ldots,N$ for prosessen, kan man gjennomføre trinnverdi iterasjonen slik:



Man starter i den øverste boksen med $v_j(0) = v_j^*(0)$ og går så gjennom syklusen for $m = 1,2,\ldots,n$. Den sekvensen $d_1^*,d_2^*,\ldots,d_n^*$ en finner er da begynnelsen på en strategi som er optimal når markovkjeden skal gjøre n overganger.

Eksempel. (Drosjeeksempel.)

Vår drosje opererer innen og mellom byene A,B,C. Hvis han er i A, (uten passasjerer), kan han forholde seg på tre måter.

1. Han kan kjøre rundt for å bli kapret på gaten.
2. Han kan kjøre til nærmeste drosjeholdeplass.
3. Han kan stoppe bilen, og vente på tur over radio.

Hvis drosjen er i C, har han de samme mulighetene. I B er det imidlertid ingen radiosentral, så der har han bare de to første mulighetene.

Under de forskjellige alternativene er det i den enkelte by forskjellige sannsynlighetsfordelinger for neste tur, og også forskjellige fortjenester.

Hvis han i A kjører rundt for å bli kapret, har han stor sannsynlighet for å få tur innen A, dessuten må han trekke utgiftene til denne kjøringen fra i fortjenesten. Likeledes for de andre tilstandene og alternativene. Under er de nødvendige data gitt.

Tilstand	Alternativ	overgangss.			fortjeneste			
i	k	P_{ij}^k			r_{ij}^k			q_i^k
		j = 1	2	3	j = 1	2	3	
1	1	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{4}$	10	4	8	8
	2	$\frac{1}{16}$	$\frac{3}{4}$	$\frac{3}{16}$	8	2	4	2.75
	3	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{5}{8}$	4	6	4	4.25
2	1	$\frac{1}{2}$	0	$\frac{1}{2}$	14	0	18	16
	2	$\frac{1}{16}$	$\frac{7}{8}$	$\frac{1}{16}$	8	16	8	15
3	1	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{2}$	10	2	8	7
	2	$\frac{1}{8}$	$\frac{3}{4}$	$\frac{1}{8}$	6	4	2	4
	3	$\frac{3}{4}$	$\frac{1}{16}$	$\frac{3}{16}$	4	0	8	4.5

Hvis drosjesjåføren har bestemt seg til at d_n skal være $\begin{bmatrix}2\\2\\1\end{bmatrix}$ ($d_n(1) = 2$, $d_n(2) = 2$, $d_n(3) = 1$) betyr det at hvis han er i A eller B etter n-te tur, skal han kjøre til nærmeste holdeplass. I C skal han kjøre rundt for å bli kapret.

La oss anta at drosjesjåføren starter med tom pung $v_i(0) = 0$, i = 1,2,3; og at han vil finne den mest lønnsomme strategien når han skal gjøre et lite antall turer.

Seks trinnverdi iterasjoner ga som resultat:

	m = 1		m = 2		m = 3	
	d^*	$v_i^*(1)$	d^*	$v_i^*(2)$	d^*	$v_i^*(3)$
i = 1	1	8	1	17.75	2	29.67
i = 2	1	16	2	29.94	2	43.42
i = 3	1	7	2	17.88	2	30.92

	m = 4		m = 5		m = 6	
	d^*	$v_i^*(4)$	d^*	$v_i^*(5)$	d^*	$v_i^*(6)$
i = 1	2	42.99	2	56.30	2	69.64
i = 2	2	56.78	2	70.13	2	83.48
i = 3	2	44.14	2	57.48	2	70.82

Vi ser at d^* har stabilisert seg på $\begin{bmatrix}2\\2\\2\end{bmatrix}$. Dette er, som vi skal se, den strategien som gir størst gevist g.

Hvis drosjen skal gjøre 4 turer bør han bruke strategien $\begin{bmatrix}2\\2\\2\end{bmatrix}, \begin{bmatrix}2\\2\\2\end{bmatrix}, \begin{bmatrix}1\\2\\2\end{bmatrix}, \begin{bmatrix}1\\1\\1\end{bmatrix}$. I begynnelsen må han tenke på framtiden og dra til holdeplass for å komme til B så ofte som mulig, derfra får han nemlig alltid gode turer.

For den siste turen behøver han bare maximere den umiddelbare fortjeneste.

Asymptotisk vurdering av strategiene.

Vi skal nå se hvordan en kan finne den strategien som er best i det lange løp. For stasjonære strategier er det før vist at den forventede fortjeneste etter n trinn $v_i(n)$ er av størrelsesorden g·n. Forventet fortjeneste

vokser over alle grenser.

Vi skal bruke den gjennomsnittelige forventede fortjeneste pr. overgang som mål på hvor god en strategi er.

For stasjonære strategier vet vi at dette gjennomsnittet er gevinsten g. Hvorledes det blir for ikke-stasjonære strategier er et mer komplisert spørsmål, men vi har følgende setning som viser at vi kan ignorere de ikke-stasjonære strategiene i jakten på den beste.

Setning 1.

Det finnes en stasjonær strategi som er minst like god som noen annen strategi.

Beviset for setningen er å finne i Annals of Mathemathical Statistics 1962, side 719, i en artikkel av Blackwell.

En måte å finne den beste strategien på er å gå gjennom alle de stasjonære strategiene, og regne ut stasjonærfordelingen og gevinsten for hver enkelt. Dette er gjort i eksemplet under.

Eksempel. (Drosjeeksempel.)

På grunnlag av de data som er gitt på side 4 , er alle de $3 \cdot 2 \cdot 3 = 18$ desisjonene undersøkt. Stasjonærfordelingen $\tilde{p}$ og gevinsten g er regnet ut for hver av dem.

$$d_9 = \begin{bmatrix} 2 \\ 1 \\ 3 \end{bmatrix} \quad P_9 = \begin{bmatrix} \frac{1}{16} & \frac{3}{4} & \frac{3}{16} \\ \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{3}{4} & \frac{1}{16} & \frac{3}{16} \end{bmatrix} \quad \tilde{p}_{d_9} = \left[\frac{200}{503}, \frac{159}{503}, \frac{144}{503}\right] \quad q_{d_9} = \begin{bmatrix} 2.75 \\ 16 \\ 4.5 \end{bmatrix}$$

$$\tilde{p}_{d_9} \cdot q_{d_9} = 6.2$$

$$d_{10} = \begin{bmatrix} 2 \\ 2 \\ 1 \end{bmatrix} \quad P_{10} = \begin{bmatrix} \frac{1}{16} & \frac{3}{4} & \frac{3}{16} \\ \frac{1}{16} & \frac{7}{8} & \frac{1}{16} \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{2} \end{bmatrix} \quad \tilde{p}_{d_{10}} = \left[\frac{2}{23}, \frac{18}{23}, \frac{3}{23}\right] \quad q_{d_{10}} = \begin{bmatrix} 2.75 \\ 15 \\ 7 \end{bmatrix}$$

$$\tilde{p}_{d_{10}} \cdot q_{d_{10}} = 12.5$$

$$d_{11} = \begin{bmatrix} 2 \\ 2 \\ 2 \end{bmatrix} \quad P_{11} = \begin{bmatrix} \frac{1}{16} & \frac{3}{4} & \frac{3}{16} \\ \frac{1}{16} & \frac{7}{8} & \frac{1}{16} \\ \frac{1}{8} & \frac{3}{4} & \frac{1}{8} \end{bmatrix} \quad \tilde{p}_{d_{11}} = \left[\frac{8}{119}, \frac{102}{119}, \frac{9}{119}\right] \quad q_{d_{11}} = \begin{bmatrix} 2.75 \\ 15 \\ 4 \end{bmatrix}$$

$$\tilde{p}_{d_{11}} \cdot q_{d_{11}} = 13.34$$

$$d_{12} = \begin{bmatrix} 2 \\ 2 \\ 3 \end{bmatrix} \quad P_{d_{12}} = \begin{bmatrix} \frac{1}{16} & \frac{3}{4} & \frac{3}{16} \\ \frac{1}{16} & \frac{7}{8} & \frac{1}{16} \\ \frac{3}{4} & \frac{1}{16} & \frac{3}{16} \end{bmatrix} \quad \tilde{p}_{d_{12}} = \left[\frac{25}{202}, \frac{159}{202}, \frac{18}{202}\right] \quad q_{d_{12}} = \begin{bmatrix} 2.75 \\ 15 \\ 4.5 \end{bmatrix}$$

$$\tilde{p}_{d_{12}} \cdot q_{d_{12}} = 12.6$$

$$d_{17} = \begin{bmatrix} 3 \\ 2 \\ 2 \end{bmatrix} \quad P = \begin{bmatrix} \frac{1}{4} & \frac{1}{8} & \frac{5}{8} \\ \frac{1}{16} & \frac{7}{8} & \frac{1}{16} \\ \frac{1}{8} & \frac{3}{4} & \frac{1}{8} \end{bmatrix} \quad \tilde{p}_{17} = \left[\frac{8}{93}, \frac{74}{93}, \frac{11}{93} \right] \quad q_{d_{17}} = \begin{bmatrix} 4.25 \\ 15 \\ 4 \end{bmatrix}$$

$$\tilde{p} \cdot q = 12.8$$

$$d_{18} = \begin{bmatrix} 3 \\ 2 \\ 3 \end{bmatrix} \quad P = \begin{bmatrix} \frac{1}{4} & \frac{1}{8} & \frac{5}{8} \\ \frac{1}{16} & \frac{7}{8} & \frac{1}{16} \\ \frac{3}{4} & \frac{1}{16} & \frac{3}{16} \end{bmatrix} \quad \tilde{p}_{18} = \left[\frac{25}{83}, \frac{36}{83}, \frac{22}{83} \right] \quad q_{d_{18}} = \begin{bmatrix} 4.25 \\ 15 \\ 4.5 \end{bmatrix}$$

$$\tilde{p} \cdot q = 9.0$$

Vi har nøyet oss med 6 av de 18 utregningene. Som man ser er det ingen liten jobb å regne ut verdien av alle strategiene. Når problemet blir større, blir det en praktisk ugjennomførbar fremgangsmåte.

I vårt problem fikk vi imidlertid løsning. Optimal-strategien er

$$d_{11} = \begin{bmatrix} 2 \\ 2 \\ 2 \end{bmatrix}.$$

VERDIDIFFERENS-ITERASJON

Vi skal nå beskrive en iterativ prosedyre som er effektiv når man skal finne den sterkeste strategien. Prosedyren går ut på at man leter seg fram i mengden D av desisjoner ved stadig å finne en som gir opphav til en bedre stasjonær strategi enn den forrige.

Vi skal som før sagt, bare ta for oss det tilfelle at for enhver desisjon har markovkjeden én rekurent klasse.

Verdi-differens-iterasjonen består av to delprosedyrer: Verdsettingsprosedyren og Forbedringsprosedyren.

Verdsettingsprosedyren bestemmer verdien av en desisjon (eller den tilhørende stasjonærstrategien) og forbedringsprosedyren finner fram til en bedre desisjon på grunnlag av disse verdiene.

Verdsettingsprosedyren.

Vi har en desisjon d. Den forventede fortjeneste etter n trinn under d når vi starter i tilstand i er $v_i(n)$. Alle størrelsene referer seg til d.

Vi har

$$v_i(n) = q_i + \sum_{j=1}^{N} p_{ij} v_j(n-1) \quad i = 1,\ldots,N \quad n = 1,2,\ldots$$

Vi har før funnet

$$v_i(n) = ng + v_i + \varepsilon_{in}$$

der ε_{in} går geometrisk mot null. De to ligningene gir

$$ng+v_i+\varepsilon_{in} = q_i + \sum_{j=1}^{N} p_{ij}\left[(n-1)g+v_j+\varepsilon_{jn-1}\right]$$

$$g+v_i = q_i + \sum_{j=1}^{N} p_{ij}v_j+\varepsilon'_{in}$$

Ved å la $n \to \infty$ får vi altså følgende N ligninger i de $N+1$ ukjente $g, v_1, \ldots, v_N$

$$(1) \qquad g+v_i = q_i + \sum_{j=1}^{N} p_{ij}v_j, \quad i = 1,2,\ldots,N.$$

Dette systemet er ubestemt, men det bestemmer g entydig. For hvis $\tilde{p} = [\tilde{p}_1,\ldots,\tilde{p}_N]$ er stasjonær-fordelingen for prosessen, får vi ved å multiplisere i-te ligning med $\tilde{p}_i$ og summere

$$g\sum\tilde{p}_i + \sum v_i\tilde{p}_i = \sum q_i\tilde{p}_i + \sum_j v_j\sum_i \tilde{p}_i \cdot p_{ij}$$

$$g = \sum_{i=1}^{N} q_i\tilde{p}_i$$

Når g er bestemt, kan (1) skrives

$$(I-P)v = q-g\cdot e$$

der $e = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}$. Nå vet vi at $I-P$ har rang $N-1$, og følgelig er den generelle løsningen $v_i = v'_i+a$, $i = 1,\ldots,N$ der v'_i, $i = 1,\ldots,N$ er en løsning av (1) og a er en

vilkårlig konstant.

Hvis vi i tillegg til (1) krever $v_N = 0$, blir systemet bestemt.

Verdsettingen av desisjonen d består i å løse ligningsystemet

$$(2) \qquad g+v_i = q_i + \sum_{j=1}^{N} p_{ij} v_j, \; v_N = 0, \; i = 1,\ldots,N.$$

g er gevinsten under desisjonen d - eller under den tilhørende stasjonære strategi. v_i er den relative fortjeneste ved å starte i tilstand i fremfor i N. Vi skal kalle v_i, $i = 1,\ldots,N$ for verdidifferensene. Vi hadde jo $v_i(n) = ng+v_i+\varepsilon_{in}$, og v_i-v_N er et uttrykk for fordelen ved å starte i tilstand i fremfor i tilstand N, når n er stor.

Forbedringsprosedyren.

Når vi har bestemt verdidifferensene v_i, $i = 1,\ldots,N$ og g for desisjonen d, skal vi bestemme en ny desisjon d' på grunnlag av verdidifferensene etter følgende skjema.

$d'(i) = k'$ hvis k' maximerer

$$q_i^k + \sum_j p_{ij}^k v_j$$

der k er et alternativ i tilstanden i. Hvis det er flere k som gir maksimum, og d(i) er blant disse skal

$d'(i) = d(i)$, hvis ikke $d(i)$ er blant de maksimerende skal $d'(i)$ være den minste av de maksimerende k. Verdi-differens-iterasjonen stopper når forbedringsprosedyren ikke lenger forandrer desisjonen, altså når

$$d'(i) = d(i) \quad i = 1,\ldots,N.$$

La P', q_i', g', v_i' og p_{ij}' for $i = 1,\ldots,N$ og $j = 1,\ldots,N$ betegne de respektive størrelsene tilhørende d'. Vi skal vise at d' er minst like god som d.

Setning 1.

d' har minst like stor gevinst som d. Hvis $d' \neq d$ og d' adskiller seg fra d for en tilstand som er rekurent under P, har d' større gevinst enn d.

Bevis. Pr. konstruksjon har vi $q_i' + \sum_j p_{ij}' v_j \geq q_i + \sum_j p_{ij} v_j$ for alle i. Med vektor-notasjon har vi altså

$$q' + P'v \geq q + Pv$$

eller $$\delta = q' - q + P'v - Pv \geq 0.$$

La e være vektoren $\begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}$. Da har vi ifølge (2)

$$ge + v = q + Pv \quad \text{og} \quad g'e + v' = q' + P'v'$$

$$(g' - g)e + v' - v = q' - q + P'v' - Pv$$

Dermed

$$(g'-g)e+v'-v = \gamma + P'(v'-v)$$

Ved multiplikasjon av linjevektoren $\tilde{p}'$ som er stasjonærfordelingen i P', gir dette

$$g'-g = \tilde{p}' \cdot \gamma \geq 0.$$

Hvis nå $d(i) = d(i)'$ for en tilstand i som er rekurrent under P', har vi $\gamma_i = q_i'-q_i + \sum_j p_{ij}' v_j - \sum_j p_{ij} v_j > 0$ og $\tilde{p}_i' > 0$. Følgelig får vi i dette tilfelle $\tilde{p}' \cdot \gamma > 0$ og $g' > g$.

Hvis verdidifferens-iterasjonen konvergerer mot d, finnes det ingen d' som har større verdi enn d.

For anta $g' > g$. Siden strategi-iterasjonen konvergerte mot d har vi

$$q'+P'v \leq q+Pv$$

og $$\gamma = q'-q+P'v-Pv \leq 0$$

Dermed får vi som i beviset over

$$g'-g = \tilde{p}' \cdot \gamma \leq 0$$

Dette er en motsigelse av $g' > g$, og følgelig eksisterer det ingen d' som er mer verdifull enn den desisjonen d som strategi-iterasjonen peker ut.

Vi har hittil vist at hvis strategi-iterasjonen konvergerer, da konvergerer den mot den mest verdifulle desisjonen. Vi har imidlertid ikke vist at strategi-iterasjonen alltid konvergerer. Vi har med andre ord ennå ikke vist om metoden fører fram.

Lemma. Hvis P har én rekurent klasse C, impliserer $Px \geq x$ at $x_i = k \quad i \in C$ og $x_j \leq k \quad j \notin C$.

Bevis. $Px \geq x \Rightarrow P^2x \geq Px \geq x \Rightarrow P^nx \geq x$

$\Rightarrow \sum_{i \in C} \tilde{p}_i x_i \geq x_j \quad j = 1,2,\ldots,N.$

Men siden $\tilde{p}_i > 0 \quad i \in C$ og $\sum_{i \in C} \tilde{p}_i = 1$ må vi ha $x_i = k \quad i \in C$. Men da er $\sum_{i \in C} \tilde{p}_i x_i = k$ og følgelig $x_j \leq k$ for $j \notin C$.

Setning 2.

Hvis det finnes en tilstand r som er rekurent under alle strategiene, konvergerer verdidifferens-iterasjonen.

Bevis. Anta at den ikke konvergerer, men peker ut sekvensen av desisjoner $d_1, d_2, \ldots$

Siden det er endelig mange desisjoner, må det finnes $n < m$ slik at $d_n = d_m$. La oss skrive g^k, q^k, v^k og $P^{(k)}$ for størrelsene tilhørende d_k. Siden $d_n = d_m$ har vi

(i) $\quad g^k = g \,, \; k = n, n+1, \ldots, m$

fordi $g^k \leq g^{k+1}$ og $g^n = g^m$

(ii) Vi må ha $v^n = v^{n+1} = \ldots = v^m$

For anta $r = N$ (det er ingen innskrenkning).

Da har vi av $g^k e + v^k = q^k + P^{(k)} v^k$

$$g^{k+1} e + v^{k+1} = q^{k+1} + P^{(k+1)} v^{k+1}$$

og $q^k + P^{(k)} v^k \leq q^{k+1} + P^{(k+1)} v^k$ (d_{k+1} ble valgt etter d_k) at

$$P^{(k+1)}(v^k - v^{k+1}) \geq (v^k - v^{k+1})$$

Av lemma har vi

$$v_i^k - v_i^{k+1} = k \text{ , } i \text{ rekurent under } P^{(k+1)}$$

$$v_j^k - v_j^{k+1} \leq k \text{ , } j \text{ ikke rekurent.}$$

Men $v_N^k = v_N^{k+1} = 0$ og N rekurent impliserer

$v_i^k = v_i^{k+1}$ når i er rekurent for $P^{(k+1)}$

$v_i^k \leq v_i^{k+1}$ når i er transient for $P^{(k+1)}$.

Dermed har vi

$$v^n \leq v^{n+1} \leq \ldots \leq v^m = v^n$$

og følgelig alle like.

Men siden verdidifferens-iterasjonen pekte ut d_{n+1} etter d_n, må den peke ut d_{n+1} etter d_{n+1} også - og vi har konvergens.

EKSEMPLER

Eksempel 1. (Drosjeeksemplet.)

Vi skal ta for oss drosjeeksemplet, og se hvorledes verdidifferens-iterasjonen virker i praksis.

De nødvendige data er gitt på side 11 .

La oss starte iterasjonen med forbedringsprosedyren på verdidifferensene $v_1 = v_2 = v_3 = 0$. Vi skal med andre ord finne de alternativene som gir størst umiddelbar forventet fortjeneste - de som maximerer q_i.

Vi finner at desisjonen $d = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$ maximerer q.

Vi skal nå bestemme verdien av d, når

$$P = \begin{bmatrix} \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{2} \end{bmatrix} \qquad q = \begin{bmatrix} 8 \\ 16 \\ 7 \end{bmatrix}$$

Systemet $g+v_i = q_i + \sum_{j=1}^{N} p_{ij} v_j \quad i = 1,\ldots,N$

$$v_N = 0$$

har løsningen $g = 9.2 \quad v_1 = 1.33 \quad v_2 = 7.47$.

Vi skal nå bruke forbedringsprosedyren på dette materialet.

Vi skal beregne $q_i^k + \sum p_{ij}^k v_j$ for alle i og k, resultat er gitt i tabellen.

i \ k	1	2	3
1	10.53*	8.43	5.52
2	16.67	21.62*	
3	9.20	9.77*	5.97

Forbedringsprosedyren gir oss desisjonen

$$d = \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}$$

Verdsettingen av denne desisjonen skjer igjen ved løsning av

$$g+v_i = q_i + \sum p_{ij} v_j, \quad v_N = 0$$

der $$P = \begin{bmatrix} \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{16} & \frac{7}{8} & \frac{1}{16} \\ \frac{1}{8} & \frac{3}{4} & \frac{1}{8} \end{bmatrix} \qquad q = \begin{bmatrix} 8 \\ 15 \\ 4 \end{bmatrix}$$

Løsningen blir $v_1 = -3.88$, $v_2 = 12.85$, $v_3 = 0$

$$g = 13.15$$

Forbedringsprosedyren.

$q_i^k + \sum_{j=1}^{N} p_{ij}^k v_j$ er gitt i tabellen

i \ k	1	2	3
1	9.27	12.14*	4.89
2	14.06	26.00*	
3	9.24	13.10*	2.39

Forbedringsprosedyren gir $d = \begin{bmatrix} 2 \\ 2 \\ 2 \end{bmatrix}$ som ny desisjon.

Verdsettingen gir ved løsning av systemet

$g+v_i = q_i + \sum p_{ij} v_j$, $v_N = 0$

der $P = \begin{bmatrix} \frac{1}{16} & \frac{3}{4} & \frac{1}{16} \\ \frac{1}{16} & \frac{7}{8} & \frac{1}{16} \\ \frac{1}{8} & \frac{3}{4} & \frac{1}{16} \end{bmatrix}$ $q = \begin{bmatrix} 2.75 \\ 15 \\ 4 \end{bmatrix}$

$v_1 = -1.18$, $v_2 = 12.66$, $v_3 = 0$, $g = 13.34$.

Forbedringsprosedyren.

$q_i^k + \sum p_{ij}^k v_j$ er gitt under

i \ k	1	2	3
1	10.58	12.17*	5.54
2	15.41	24.42*	
3	9.87	13.34*	4.41

Forbedringsprosedyren gir oss igjen desisjon $d = \begin{bmatrix} 2 \\ 2 \\ 2 \end{bmatrix}$, som dermed er det endelige resultatet av verdidifferens-iterasjonen.

Drosjen bør altså - som vi før har funnet - alltid dra til nærmeste holdeplass. Med en slik strategi vil han gjennomsnittelig tjene 13.34 pr. tur.

Når vi regner med alle gangene, har vi brukt forbedringsprosedyren 4 og verdsettingsprosedyren 3 ganger.

En sammenligning av det arbeidet som skulle til for å gjennomføre denne verdidifferens -iterasjonen, og det arbeidet som skulle til for å finne $\tilde{p}_d \cdot q_d$ for alle desisjoner d, faller klart ut til fordel for verdi-differens -iterasjonen.

Når antall tilstander og antall alternativer er større, blir strategi-iterasjonen forholdsvis mye mer effektiv enn den direkte metoden.

Eksempel 2.

Dette eksemplet er typisk for mange problemer som kan løses ved dynamisk styring. Det typiske er at man har en produksjonsenhet som foreldes. Problemet er når det lønner seg å selge den enheten man har, og hvor gammel skal den være som man kjøper ?

I eksemplet gjelder det fornyelse av bil. Hvert kvartal vurderer vi om vi skal selge bilen, og i tilfelle salg hvor gammel bil vi skal kjøpe istedet. Markovkjeden $\langle X_n \rangle$ er definert ved $X_n = j$ når den bilen vi har på tidspunkt n er j perioder à 3 måneder gammel. For å holde antall tilstander begrenset, skal j løpe fra 1,...,40. Tilstand 40 svarer til at vår bil er 10 år gammel

eller eldre, eller at den har brutt sammen og er ubrukbar.

Ved utgangen av hvert kvartal kan vi på grunnlag av X_n velge alternativ k. Alternativet $k = 1$ er å beholde bilen. Alternativet $k > 1$ er å selge bilen, og kjøpe en ny som er $k-2$ perioder gammel. $k = 1,2,\ldots,41$.

Det er 41^{40} forskjellige desisjonsfunksjoner.

For oss skal følgende størrelser være tilstrekkelige for å beskrive en i-perioder gammel bils økonomi.

C_i = innkjøpspris for en bil av alder i.

T_i = salgsprisen for en bil av alder i.

E_i = de forventede driftsutgifter i en periode for en bil av alder i.

p_i = sannsynligheten for at en i perioder gammel bil vil holde en periode til uten sammenbrudd.

Vi skal skrive ned relasjonen

$$g+v_i = q_i^k + \sum_{j=1}^{40} p_{ij}^k v_j \quad \text{for} \quad k = 1,\ldots,41$$

q_i^1 = umiddelbar fortjeneste i en periode = $-E_i$.

q_i^k = -kostnadene for å bytte bil pluss den nyes drifts-utgifter = $T_i - C_{k-2} - E_{k-2}$ $\qquad k = 2,\ldots,41$

$$p_{ij}^1 = \begin{cases} p_i & j = i+1 \\ 1-p_i & j = 40 \\ 0 & \text{ellers} \end{cases} \qquad p_{ij}^k = \begin{cases} p_{k-2} & j = k-1 \\ 1-p_{k-2} & j = 40 \\ 0 & \text{ellers} \end{cases}$$

$k = 2,\ldots,41.$

$g+v_i = q_i^k + \sum p_{ij}^k v_j$ blir

for $k = 1$: $g+v_i = -E_i + p_i v_{i+1} + (1-p_i) v_{40}$,

for $k \geq 2$: $g+v_i = T_i - C_{k-2} - E_{k-2} + p_{k-2} v_{k-1} + (1-p_{k-2}) v_{40}$

som gjelder for $i = 1,\ldots,40$.

Dette eksemplet er hentet fra boka til Howard, og de tallene som er brukt for å finne den beste strategien passer nok bedre i U.S.A. enn her.

Data for bilfornyelses-eksempelet

i	C_i	T_i	E_i	p_i	i	C_i	T_i	E_i	p_i
0	$2000	$1600	$50	1.000					
1	1840	1460	53	0.999	21	$345	$240	$115	0.925
2	1680	1340	56	0.998	22	330	225	118	0.919
3	1560	1230	59	0.997	23	315	210	121	0.910
4	1300	1050	62	0.996	24	300	200	125	0.900
5	1220	980	65	0.994	25	290	190	129	0.890
6	1150	910	68	0.991	26	280	180	133	0.880
7	1080	840	71	0.988	27	265	170	137	0.865
8	900	710	75	0.985	28	250	160	141	0.850
9	840	650	78	0.983	29	240	150	145	0.820
10	780	600	81	0.980	30	230	145	150	0.790
11	730	550	84	0.975	31	220	140	155	0.760
12	600	480	87	0.970	32	210	135	160	0.730
13	560	430	90	0.965	33	200	130	167	0.660
14	520	390	93	0.960	34	190	120	175	0.590
15	480	360	96	0.955	35	180	115	182	0.510
16	440	330	100	0.950	36	170	110	190	0.430
17	420	310	103	0.945	37	160	105	205	0.300
18	400	290	106	0.940	38	150	95	220	0.200
19	380	270	109	0.935	39	140	87	235	0.100
20	360	255	112	0.930	40	130	80	250	0

Bilfornyelsesproblemet ble løst i 7 iterasjoner. Den optimale strategien ble funnet å være : hvis den bilen en har er mellom $\frac{1}{2}$ og $6\frac{1}{2}$ år bør man beholde den. Hvis bilen er yngre enn $\frac{1}{2}$ år eller eldre enn $6\frac{1}{2}$ år, bør den byttes ut med en bil som er 3 år gammel. Denne strategien har gevinst - 150.95 dollar. De forventede utgiftene til transport er altså 150.95 dollars pr. år. Vi ser at verdien øket sterkest ved de første iterasjonene. Ved de 3-4 siste strategiene, øket ikke verdien mye. Iterasjonen foretok der bare en finpuss av den endelige strategien.

I resultatmatrisen er oppgitt verdidifferensene v_i. v_i angir fordelen ved å starte i tilstand i fremfor i tilstand 40. I siste kolonne står den justerte verdi. Der er v_i skalert ved at $v_{40} = 80$, som er salgsprisen for en 10 år gammel bil. De justerte v_i betegner da i dollar hvor verdifull en i perioder gammel bil er for en bruker.

Hvis jeg har en 5 år gammel bil, og blir tilbudt en 2 år gammel bil, vil jeg bare kjøpe den hvis prisen for den 2 år gamle bilen er mindre eller lik $v_8 - v_{20} = 840 - 295 = 545$ dollar.

I resultatmatrisen er desisjonen 1 byttet ut med K.

Resultatet av 7 iterasjoner

Tilstand = i	Iterasjon 1 g = -250.00 d(i)	v_i	Iterasjon 2 g = -193.89 d(i)	v_i	Iterasjon 3 g = -162.44 d(i)	v_i	Iterasjon 4 g = -157.07 d(i)	v_i
1	36	$1374	20	$1380	19	$1380	12	$1380
2	36	1254	20	1260	19	1260	12	1260
3	36	1144	20	1150	19	1150	12	1150
4	36	964	20	970	K	1037	12	970
5	36	894	20	900	K	940	12	900
6	36	824	20	830	K	848	12	830
7	36	754	20	760	19	760	12	760
8	36	624	20	630	K	696	12	630
9	36	564	20	570	K	617	12	570
10	36	514	20	520	K	542	12	520
11	36	464	20	470	19	470	12	470
12	36	394	20	400	19	400	K	520
13	36	344	20	350	K	575	K	464
14	36	304	20	310	K	521	K	411
15	36	274	20	280	K	470	K	362
16	36	244	20	250	K	423	K	315
17	36	224	20	230	K	380	K	271
18	36	204	20	210	K	338	K	230
19	36	184	20	190	K	300	12	190
20	36	169	K	280	K	264	12	175
21	K	876	K	213	K	229	12	160
22	K	801	20	145	K	197	12	145
23	K	728	20	130	K	166	12	130
24	K	658	20	120	K	136	12	120
25	K	592	20	110	19	110	12	110
26	K	530	20	100	19	100	12	100
27	K	469	20	90	19	90	12	90
28	K	412	20	80	19	80	12	80
29	K	356	20	70	19	70	12	70
30	K	306	20	65	19	65	12	65
31	K	261	20	60	19	60	12	60
32	K	218	20	55	19	55	12	55
33	K	176	20	50	19	50	12	50
34	K	140	20	40	19	40	12	40
35	K	111	20	35	19	35	12	35
36	K	84	20	30	19	30	12	30
37	K	55	20	25	19	25	12	25
38	K	33	20	15	19	15	12	15
39	K	15	20	7	19	7	12	7
40	K	0	20	0	19	0	12	0

Tilstand = i	Iterasjon 5 g = -151.05 d(i)	v_i	Iterasjon 6 g = -150.99 d(i)	v_i	Iterasjon 7 g = -150.95 d(i)	v_i	justert v(i)
1	12	$1380	12	$1380	12	$1380	$1460
2	12	1260	12	1260	12	1260	1340
3	12	1150	12	1150	K	1161	1241
4	K	1003	K	1072	K	1072	1152
5	K	917	K	987	K	987	1067
6	K	836	K	907	K	906	986
7	12	760	K	831	K	831	911
8	K	761	K	760	K	760	840
9	K	695	K	695	K	695	775
10	K	633	K	633	K	632	712
11	K	574	K	574	K	574	654
12	K	520	K	520	K	520	600
13	K	470	K	470	K	470	550
14	K	424	K	424	K	424	504
15	K	381	K	381	K	381	461
16	K	341	K	342	K	342	422
17	K	306	K	306	K	306	386
18	K	273	K	273	K	273	353
19	K	242	K	243	K	243	323
20	K	214	K	214	K	215	295
21	K	188	K	189	K	189	269
22	K	164	K	165	K	166	246
23	K	143	K	144	K	144	224
24	K	124	K	125	K	126	206
25	K	109	12	110	K	111	191
26	K	97	12	100	12	100	180
27	12	90	12	90	12	90	170
28	12	80	12	80	12	80	160
29	12	70	12	70	12	70	150
30	12	65	12	65	12	65	145
31	12	60	12	60	12	60	140
32	12	55	12	55	12	55	135
33	12	50	12	50	12	50	130
34	12	40	12	40	12	40	120
35	12	35	12	35	12	35	115
36	12	30	12	30	12	30	110
37	12	25	12	25	12	25	105
38	12	15	12	15	12	15	95
39	12	7	12	7	12	7	87
40	12	0	12	0	12	0	80

II KONTINUERLIG TID

Vi har en situasjon som kan beskrives som en markov-prosess med kontinuerlig tid, med et endelig antall tilstander $1,2,\ldots,N$ og med intensitetsmatrise $A = \{a_{ij}\}$. Til prosessen er det knyttet en nyttefunksjon som i det diskrete tilfelle. Ved overgang fra tilstand i til tilstand j, tjener prosessen r_{ij}, og den tjener r_{ii} pr. tidsenhet når den er i tilstand i.

$v_i(t)$ er den forventede fortjeneste i tidsrommet $(0,t)$ når $X(0) = i$. Når $V_i(t)$ er den stokastiske variable fortjeneste i tidsrommet $(0,t)$ gitt $X(0) = i$, har vi

$$\begin{aligned} v_i(t+dt) &= EV_i(t+dt) = EE\left[V_i(t+dt)\,\middle|\,X(dt)\right] \\ &= (1+a_{ii}dt)E\left[V_i(t+dt)\,\middle|\,X(dt)=i\right] + \\ &\qquad + \sum_{j\neq i} a_{ij}dtE\left[V_i(t+dt)\,\middle|\,X(dt)=j\right] \\ &= (1+a_{ii}dt)\cdot(r_{ii}dt+v_i(t)) + \sum_{j\neq i} a_{ij}dt(r_{ij}+v_j(t)) \end{aligned}$$

Dette gir relasjonen

$$v_i'(t) = r_{ii} + \sum_{j\neq i} a_{ij}r_{ij} + \sum_{j=1}^{N} a_{ij}v_j(t)$$

Ved å sette $q_i = r_{ii} + \sum_{j\neq i} a_{ij}r_{ij}$, $q = \begin{bmatrix} q_1 \\ \vdots \\ q_N \end{bmatrix}$ og $v(t) = \begin{bmatrix} v_1(t) \\ \vdots \\ v_N(t) \end{bmatrix}$

får vi

$$(1) \qquad \frac{\partial}{\partial t}v(t) = q+A\cdot v(t).$$

Hvis nå $\varphi_i(s)$ er den Laplace-transformerte av $v_i(t)$, og $\Phi(s)$ den Laplacetransformerte av $v(t)$, får vi av (1)

$$s\cdot \Phi(s)-v(0) = \frac{1}{s}\cdot q+A\cdot \Phi(s)$$

$$(sI-A)\cdot\Phi(s) = \frac{1}{s}q+v(0)$$

$$\Phi(s) = (sI-A)^{-1}(\frac{1}{s}q+v(0))$$

Som bemerket i "Operatormetoder", kan $(sI-A)^{-1}$ spaltes opp i en stasjonær del og en transient del når prosessen har én rekurent klasse.

$$(sI-A)^{-1} = \frac{1}{s}\cdot S+\Gamma(s)$$

Når alle røttene i $|sI-A|$ er forskjellige har vi $\Gamma(s) = \sum_{i=1}^{N-1} \frac{1}{s_i+s} T_i \quad s_i > 0.$

I dette tilfellet får vi

$$\Phi(s) = \left[\frac{1}{s}S+\Gamma(s)\right]\left[\frac{1}{s}\cdot q+v(0)\right]$$

$$= \frac{1}{s^2}Sq+\frac{1}{s}\Gamma(s)\cdot q+\frac{1}{s}Sv(0)+\Gamma(s)v(0)$$

Den Laplacetransformerte til en sum er summen av de Laplacetransformerte. Vi skal derfor undersøke hvert ledd i summen over.

$$\frac{1}{s^2}\mathcal{S}q = \int_0^\infty te^{-st}dt \cdot \mathcal{S}q$$

Når nullpunktene $-s_i$ er forskjellige, har vi

$$\frac{1}{s}\mathcal{T}(s) = \sum_{i=1}^{N-1} \frac{1}{s}\frac{1}{s+s_i}\cdot T_i = \sum_{i=1}^{N-1} T_i\left(\frac{1}{s_i}\cdot\frac{1}{s} - \frac{1}{s_i}\cdot\frac{1}{s+s_i}\right)$$

$$= \sum T_i \cdot \frac{1}{s_i}\left[\int_0^\infty e^{-st}dt - \int_0^\infty e^{-s_i t}\cdot e^{-st}dt\right] =$$

$$= \int_0^\infty \mathcal{T}(0)e^{-st}dt + \int_0^\infty \mathcal{E}_1(t)e^{-st}dt$$

Der $\mathcal{E}_1(t)$ går eksponensielt mot null.

$$\frac{1}{s}\mathcal{S}v(0) = \int_0^\infty \mathcal{S}v(0)\cdot e^{-st}dt \quad \text{og} \quad \mathcal{T}(s)v(0) = \int_0^\infty \mathcal{E}_2(t)e^{-st}dt$$

$\mathcal{E}_2(t)$ går også eksponensielt mot null.

Vi har altså

$$\bar{\Phi}(s) = \int_0^\infty \left[t\cdot\mathcal{S}q + \mathcal{T}(0)\cdot q + \mathcal{S}v(0) + \mathcal{E}(t)\right]\cdot e^{-st}dt$$

følgelig

$$v(t) = t\cdot\mathcal{S}q + \mathcal{T}(0)q + \mathcal{S}v(0) + \mathcal{E}(t)$$

Denne formen har $v(t)$ også om ikke alle nullpunktene er forskjellige.

Vi skal forutsette at prosessen bare har én rekurent klasse. Da kan vi skrive

$\mathcal{S}\cdot q = g\cdot e$ der g er gevinsten for denne prosessen og $e = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}$. $\mathcal{T}(0)\cdot q + \mathcal{S}v(0) = v = \begin{bmatrix} v_1 \\ \vdots \\ v_N \end{bmatrix}$.

$v_i - v_j$ er fordelen ved å starte i tilstand i fremfor i j.

Vi har følgelig

$$v(t) = t\cdot g\cdot e + v + \mathcal{E}(t) \tag{2}$$

La oss nå definere desisjonsproblemet.

Vi har tilstandene $1,2,\ldots,N$. I tilstand i kan vi velge mellom n_i alternativer. Hvert alternativ k består av en intensitetsvektor a_i^k og en fortjenestevektor r_i^k. Som i det diskrete tilfellet lar vi $D = \{d \mid d(i) \in \{1,2,\ldots,n_i\}\}$ være mengden av desisjonene. Hvis $X(t) = i$ og $d(i) = k$ vil prosessen bevege seg i henhold til a_i^k og tjene i henhold til r_i^k intil den gjør en overgang.

En strategi $\delta = \langle d_1, d_2, \ldots \rangle$ virker slik: hvis n-te overgang skjer på tidspunktet t_n og $(n+1)$-te overgang på t_{n+1}, $X(t) = i$ $t_n < t \leq t_{n+1}$ vil $X(t)$ være behersket av intensitetsvektoren $a_i^{d_n(i)}$ og tjene i henhold til $r_i^{d_n(i)}$.

Eksempel 1. (Maskineksempel, se "Operatormetoder" side 43).

Vår maskin kan være i tilstandene

1 - gå for full fart

2 - gå for halv fart (være delvis i stykker)

3 - være ute av drift (helt i stykker).

I tilstanden $i = 1$ har vi alternativene billig vedlikehold $k = 1$, dyrt vedlikehold $k = 2$.

I tilstanden $i = 2$ har vi mulighetene

$k = 1$: bruke husets reparatør - det er billigst, men han er ikke så flink

$k = 2$: bruke spesialist - dyrt men godt.

I tilstand $i = 3$ har vi mulighetene

$k = 1$: bruke husets reparatør

$k = 2$: bruke spesialist

$k = 3$: skifte maskin.

Data skjema:

Tilstand	Alternativ	Overgangsintensiteter			Fortjenester			
i	k	a_{i1}^k	a_{i2}^k	a_{i3}^k	r_{i1}^k	r_{i2}^k	r_{i3}^k	q_i^k
1	1	-3	1	2	22	0	0	22
	2	-1	0.667	0.333	20	0	0	20
2	1	3	-4	1	0	2	0	2
	2	7	-7.5	0.5	-2	1	-2	-14
3	1	1	3	-4	0	0	-4	-4
	2	3	0.5	-3.5	-2	-2	-5	-12
	3	20	0	-20	-100	-100	-4	-2004

I tilstand 3, alternativ 3 har vi

$a_i^k = [20 \quad 0 \quad -20]$. Det kommer av at den nye maskinen har eksponentielt fordelt innkjøringstid. En ny maskin koster 100. At $r_2^2 = [-2, 1, -2]$ kommer av at spesialisten krever 2 i fast avgift + timebetaling. Strategien $\begin{bmatrix} 2 \\ 1 \\ 2 \end{bmatrix}$ går ut på:

Når maskinen går skal den ha dyrt vedlikehold. Når den går halvveis i stykker skal den repareres av egen reparatør. Når den går helt i stykker skal den repareres av spesialist.

VERDIDIFFERENS-ITERASJON

Den forventede fortjeneste pr. tidsenhet er det mål vi skal bruke på hvor god en strategi er. For en stasjonær strategi $\delta = \langle d \rangle$ har vi av formelen

$$v^d(t) = t \cdot g^d e + v^d + \varepsilon^d(t)$$

at $g^d = S^d q^d$ er et mål på denne strategien.

Som i det diskrete tilfellet, skal vi begrense oss til å finne den beste strategien blant de stasjonære strategiene. Vi skal med andre ord finne den desisjonen d som gir opphav til den markovprosessen som har størst gevinst g^d.

Vi skal presentere verdidifferens- metoden som også i dette tilfellet er effektiv når det gjelder å finne den beste stasjonære strategien.

Verdsettingsprosedyren.

Ved å innsette $v(t) = tge + v + \varepsilon(t)$ i $\frac{\partial}{\partial t} v(t) = q + A \cdot v(t)$ får vi

$$ge + \varepsilon'(t) = q + tAg \cdot e + Av + A\varepsilon(t)$$

Nå er $\varepsilon(t)$ et eksponensielt polynom som går mot null, og følgelig går $\varepsilon'(t)$ også mot null når

$t \to \infty$. Siden $\sum_{j=1}^{N} a_{ij} = 0$, er $A \cdot g \cdot e = g \cdot A \cdot e = 0$.

Ved å la $t \to \infty$ må vi altså ha

$$g \cdot e = q + Av. \tag{3}$$

Dette er et system av N ligninger i de N+1 ukjente $g, v_1, \ldots, v_N$, men siden det bare er én rekurent klasse, er

$$\text{rang } A = N-1.$$

Nå vet vi imidlertid at $\tilde{p} \cdot A = 0$, $\tilde{p} = \lim_{t \to \infty} P(t)$, og dermed impliserer (3) $g = \tilde{p} \cdot q$. Når g er bestemt, blir v bestemt på en additiv konstant nær. Ved å kreve $v_N = 0$ får vi altså et bestemt system.

Verdsettingsprosedyren består i å løse systemet

$$ge = q + Av, \quad v_N = 0.$$

Forbedringsprosedyren.

Hvis vi skal bruke en strategi t tidsenheter framover, og så finner ut at vi har mere tid som vi vil utnytte godt, ser vi av

$$\frac{\partial}{\partial t} v(t) = q + Av(t)$$

at det må være lurt å maksimere $\frac{\partial}{\partial t} v(t)$, ved å finne den d som maksimerer $q^d + A^d \cdot v(t)$.

For hver i skal vi finne den k som maksimerer $q_i^k + \sum_{j=1}^{N} a_{ij}^k \cdot v_j(t)$.

Nå har vi $v_j(t) = t \cdot g + v_j + \varepsilon_j(t)$, der $\varepsilon_j(t)$ kan neglisjeres.

Vi skal altså maksimere $q_i^k + \sum_{j=1}^{N} a_{ij}^k v_j$.

Forbedringsprosedyren består nå i, på grunnlag av en desisjon d med tilhørende g^d og v_j^d, å finne desisjonen d' slik at $k = d'(i)$ maksimerer $q_i^k + \sum_{j=1}^{N} a_{ij}^k v_j$ for $i = 1,\ldots,N$.

Som i det diskrete tilfellet skal vi vise at verdi-differens -iterasjonen leder oss til stadig bedre desisjoner.

Setning.

Hvis d' er bestemt på grunnlag av d er $g' \geq g$.

Bevis. Vi skal sette merke på alle størrelsene som refererer seg til d'. Pr. konstruksjon har vi

$$q' + A'v \geq q + Av$$

$$\gamma = q' - q + A'v - Av \geq 0$$

Nå er $g \cdot e = q + Av$ og $g'e = q' + A'v'$ følgelig

$$(g' - g) \cdot e = q' - q + A'v' - Av + A'v - A'v$$
$$= \gamma + A'(v' - v)$$

Men $\tilde{p}' \geq 0$ og $\tilde{p}'A = 0$; dermed

$$g'-g = \tilde{p}' \cdot \gamma \geq 0.$$

På samme vis som i det diskrete tilfellet, ser en at hvis verdidifferens-iterasjonen konvergerer, så konvergerer den mot den beste strategien.

At verdidifferens-iterasjonen virkelig konvergerer når det er en tilstand som er rekurent under alle desisjonene, bevises også som i det diskrete tilfellet ved å se på P der

$$p_{ij} = \begin{cases} -\frac{a_{ij}}{a_{ii}} & i \neq j \\ 0 & i = j \end{cases} .$$

Eksempel 2.

Vi skal finne den strategien som er best i det lange løp ved verdidifferens-iterasjon i maskineksemplet.

Vi skal starte med den desisjonen som gir best umiddelbar fortjeneste. Det er $d = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$ med $q = \begin{bmatrix} 22 \\ 2 \\ -4 \end{bmatrix}$

Verdsettingen.

Systemet

$$g = 22-3v_1+v_2+2v_3$$
$$g = 2+3v_1-4v_2+v_3$$
$$g = -4+v_1+3v_2-4v_3 \quad \text{og} \quad v_3 = 0$$

gir $g = 8.45 \quad v_1 = 5.31 \quad v_2 = 2.38.$

Forbedring.

$q_i^k + \sum_j a_{ij}^k v_j$ er tabellert under:

i \ k	1	2	3
1	8.45	16.277	
2	8.45	5.320	
3	8.45	5.120	-1897.8

som gir $d' = \begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix}$.

Verdsettingen gir

$g = 13.891 \qquad v_1 = 8.249 \qquad v_2 = 3.214$

Forbedring.

$q_i^k + \sum_j a_{ij}^k v_j$

i \ k	1	2	3
1	0.467	13.891	
2	13.891	19.638	
3	13.891	14.354	-1839.12

$d' = \begin{bmatrix} 2 \\ 2 \\ 2 \end{bmatrix}$

Verdsetting.

$g = 14.395 \qquad v_1 = 8.16 \qquad v_2 = 3.83$

Forbedring.

$q_i^k + \sum_j a_{ij}^k v_j$

i \ k	1	2	3
1	1.35	14.395	
2	11.16	14.395	
3	15.65	14.395	-1840.8

$$d' = \begin{bmatrix} 2 \\ 2 \\ 1 \end{bmatrix}$$

Verdsetting gir

$$g = 14.496 \qquad v_1 = 7.867 \qquad v_2 = 3.543$$

Forbedring.

$q_i^k + \sum_j a_{ij}^k v_j$

i \ k	1	2	3
1	1.942	14.496	
2	11.429	14.496	
3	14.496	13.372	stor negativ

$$d' = \begin{bmatrix} 2 \\ 2 \\ 1 \end{bmatrix}$$

Vi har konvergens. Den beste strategien i det lange løp er dyrt vedlikehold, bruk spesialist når den går for halv fart og bruk husets reparatør når maskinen er helt i stykker.

Gevinsten pr. tidsenhet er 14.5 under denne strategien.